White Paper Report

Report ID: 98810

Application Number: PR5008010

Project Director: Daniel Pitti (dpitti@virginia.edu)

Institution: University of Virginia

Reporting Period: 5/1/2010-4/30/2013

Report Due: 7/31/2013

Date Submitted: 8/1/2013

Report type: White Paper

Grant number: PR-50080-10

Title: Enhancing Access to Contextual Information on Individuals, Families, and Corporate Bodies for Archival Collections (Social Networks and Archival Context: SNAC)

Project director: Daniel V. Pitti

Name of grantee institution: University of Virginia

Date report submitted: July 31, 2013

The Social Networks and Archival Context Project (SNAC)

Final Performance Report, covering May 1, 2010, through April 30, 2013

Project Activities

The Social Networks and Archival Context (SNAC) began May 1, 2010, and concluded on April 30, 2013. The project was initially scheduled to be completed April 30, 2012, but was extended to accommodate a supplement to the grant, to fund revisions of Encoded Archival Description (EAD) incorporating recommendations from the project. The research and development work was completed in April 2012. In 2012 the Andrew W. Mellon Foundation, based on the work completed, funded SNAC for an additional two years, expanding the research agenda, vastly increasing the quantity and diversity of the source data, and developing a foundation for establishing a national archival description cooperative.

The project staff were located at three institutions: Institute for Advanced Technology in the Humanities, University of Virginia (IATH: lead institution); School of Information, University of California, Berkeley (SI/UCB); and the California Digital Library, University of California (CDL). The project had the following staff: Daniel Pitti (Project Director) and Worthy Martin (Co-PD) at IATH; Ray Larson and Krishna Janakiraman (graduate student) at SI/UCB; Brian Tingle and Adrian Turner at CDL. In addition, Sarah Wells and Shayne Brandon (IATH), and Ivy Anderson and Sherri Berger (CDL) have made contributions to the project.

SNAC's broad objective is to contribute to the ongoing transformation of archival description, using advanced technologies in order to improve its quality and efficiency and thereby enhancing user access to and understanding of humanities primary resources. In this regard, this phase of SNAC specifically focused on exploring the feasibility of separating the description of people from the description of the resources that document their lives.

Central to the project is the observation that the predominate form of archival description is a single, monolithic description of a collection of records having the same creator (common provenance). These descriptions, commonly known as finding aids, combine descriptions of the records as such as well as descriptions of the creator of the records and, frequently, others that are documented in the records (context). SNAC envisions that future archival description systems will maintain the descriptions of records and the descriptions of people documented in the records separately though interrelated. This will improve the efficiency and quality of archival description and make it possible to provide users not only with the traditional form of access (the finding aid), but also with additional pathways into humanities resources.

In order to test the feasibility of this vision of archival description, SNAC focused on the following three activities: 1) extracting data describing people (names and historical information) from finding aids and assembling the extracted data into archival authority records (using the archival authority record standard Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF)); 2) matching the authority records with one another as well as with library authority records, and combining the data from matching records; and 3) using the resulting unique set of records to develop a prototype public system that provides integrated access to dispersed archival records and, at the same time, a resource

documenting the socio-historical context of the people found in the descriptions. The three activities were distributed among the project partners: extraction/assembling at IATH; matching/merging at the SI/UCB; and developing the prototype at CDL.

The Library of Congress (LoC) and three consortia — the Online Archive of California (OAC), the Northwest Digital Archive (NWDA), and Virginia Heritage (VH) — provided the archival description source data for the project, EAD-encoded finding aids. OCLC contributed the Virtual International Authority File (VIAF) records, the Library of Congress the NACO/LCNAF, and the Getty Research Institute the Union List of Artist Names (ULAN). The LCNAF and ULAN were not directly used in SNAC because both files are available in VIAF.

Over the course of the project, project collaborators have given over twenty public presentations. Among the presentations venues are the following: OCLC VIAF Council meeting (Strasbourg); Archives nationales (Paris); APEx (Bundesarchiv, Berlin); CNI (two, Washington, D.C.); RBMS (San Diego); National Archives and Records Administration (Washington, D.C.); Digital Humanities Conference (2: Palo Alto and Hamburg); ICA Congress (Brisbane); SAA (two, Chicago and San Diego); Theory and Practice of Digital Libraries (Berlin); and DLF (two, Palo Alto and Denver).

Ray Larson's and Krishna Janakiraman's "Connecting Archival Collections: The Social Networks and Archival Context Project" was cited as the best paper at the TPDL conference in Berlin in 2011. The project was also featured in a Chronicle of Higher Education Article: "Projects Aims to Build Online Hub for Archival Materials," http://chronicle.com/article/Building-a-Digital-Map-of/131846/

In the summer of 2011, IATH submitted the proposal "Building a National Archival Authorities Infrastructure" to the Institute of Museum and Library Services (IMLS), which was funded for the period October 2011 through September 2013. This project, currently underway, is engaged in two activities that are leveraging and building on the SNAC project: teaching SAA EAC-CPF workshops regionally and developing a blueprint for an archival access cooperative.

Also in 2011, IATH applied to NEH for a small SNAC supplemental grant to fund meetings associated with the revision of the archival standard Encoded Archival Description (EAD). The revision, in part, has focused on changes that will improve the interoperation of EAD with Encoded Archival Context-Corporate Bodies, Persons, Families (EAC-CPF), the underlying standard employed in the SNAC project.

Accomplishments

In the initial NEH-funded phase, SNAC had four core objectives:

- Support scholars and other users in discovering and identifying persons, families, and organizations by making all of the names used by and for them searchable;
- Enhance access to primary resources by linking descriptions of people to a wide variety of resources by and about them;
- Provide access to the social and profession networks within which people live and work by systematically documenting their relationships with one another;

- Provide archives with an efficient and reliable means to exploit and transform archival description, thereby improving description of archival records and re-using metadata across repositories.

Each of these four objectives was intended to serve the overarching goal of "transforming description of and improving access to primary sources through the use of advanced technologies."

In order to achieve these objectives, SNAC engaged in a sequence of three activities:

1. Developing software to extract names and biographical/historical data from EAD-encoded finding aids and using the data to assemble into Encoded Archival Context-Corporate Bodies, Persons, and Families (EAD-CPF) records (henceforth CPF records).
2. Matching the CPF records first against one another and then against authority records in OCLC's Virtual International Authority File (VIAF), combining data from matching CPF records, and augmenting the descriptive information with data from matching VIAF records.
3. Based on the set of records produced by the first two steps, developing a prototype access system and historical resources.

The first step, extracting CPF authority records from EAD-encoded finding aids, was done at IATH. The second step, matching and combining the CPF records with one another and with authority records, was done at SI/UCB. The third step, developing the prototype public system, was performed at CDL.

Extracting EAD Data /Assembling EAC-CPF Records - IATH

IATH acquired 30,109 EAD-encoded findings from LoC, NWDA, OAC, and VH. A program was developed using the languages EXtensible Stylesheet Language Transformations (XSLT) and XPath to extract data from the finding aids and assemble the data into CPF records. The program focuses on extracting names that have been tagged as person name, corporate name, or family name. The role of the named entity in relation to the described records is determined by a number of contextual factors. Creators of the described record are identified by the occurrence of the tagged name in the <origination> element in the finding aid. All other names found are simply designated as "associated with" the records, and by extension, with the creator of the records, suggesting that there was (or is) a *social*, *professional*, or *intellectual* relation between the creator and the named entity. Regular Expressions were used to analyze name strings for the presence of some variation of the word "correspondence" in order to identify entities with whom the creator corresponded. Names found in the detailed description ("description of subordinate components") in the finding aid were also designated as "associate with," though XPATH and Regular Expressions were used to identify when names occurred within the description of correspondence.

For the creator of the described records, the program also extracted biographical/historical information from the finding aid. Such information is in the form of chronological lists (or timelines) that document the life of the creator. The program also extracted occupation data, as well as subject entries used in describing the records. While the latter reflect the intellectual content of the described records, the fact

that the creator is associated with records addressing specific subjects is helpful in identifying and understanding the identity of the named entity.

As the data is being extracted from the finding aids, the relations among the named entities and associated data are maintained for use in assembling the CPF records. For each name, the finding aid source is maintained. For creators, the biographical/historical data, occupations, associated subjects, and the names of all other persons, corporate bodies, and families are also maintained. For referenced names (names other than the creator), the name of the creator the records is maintained. Maintaining the related data makes it possible to assemble CPF records for each unique name found, and to link the named entity to the source finding aid, as well as to related named entities. The links to the finding aids, once CPF records have been matched/merged, enable providing integrated access to the described resources and the links to other named entities enable exposing the social-professional-intellectual relations of each named entity.

Extracting data from the finding aids and assembling the CPF records presents a number of challenges, some of which can be overcome and others only ameliorated. Some archival collections have multiple creators. When such occurs, it is difficult to associate any biography/history information with the correct entity because the EAD-encoding does not relate the information unambiguously. In addition, some finding aids have used the biographical/history section of the finding aid to incorrectly create timelines that are not, in fact, biographies or histories of the named entity. While it is possible to programmatically associate the correct biography or history with a reasonable amount of accuracy, some incorrect associations still occur.

Further challenges to extraction and assembling processing are presented by inconsistencies in tagging and formulation of names. There are multiple instances of person names identified as corporate names, corporate names as person names, family names as corporate or person names. Using keyword lists, it is possible to identify some errors in the tagging of corporate bodies and families. There are also many examples of the names of multiple people enclosed in one tag, when In fact each name should be enclosed in a separate tag. Personal names are sometimes in direct order (forename followed by surname) and sometimes indirect (surname followed by forename), and again, while this can be ameliorated, many inconsistencies remain in the output. Finally, the punctuation and casing of names is inconsistent. While one may be tempted to be critical of the archival community with respect to the uneven quality and inconsistent practices, no one anticipated at the time the finding aids were created that the data would be used in the manner that SNAC is using it.

While the intent was to release the extraction/assembling software as open source at the end of the project period, it was decided that the software, while largely successful in producing CPF records of an acceptable quality, was overly complex, rendering it difficult to use by others without substantial support, and furthermore was difficult to maintain. Thus we have deferred the release of the software as open source until it can be thoroughly rewritten, informed by the knowledge and experience acquired in the initial development. This thorough revision is nearing completion in the second Mellon-funded phase of SNAC.

The following table provides a quantitative overview of the extraction/assembling step of the processing.

Quantitative overview of extraction and assembling

Repository	*Finding Aids*	*Person*	*Corporate*	*Family*	*Total*
LOC	1,159	36,012	7,243	447	43,702
NWDA	5,160	13,294	10,303	1,355	24,952
OAC	~15,400	66,330	24,860	624	91,814
VH	8,390	9,919	4,783	473	15,175
Total	30,109	125,555	47,189	2,899	175,643

The process for extracting data and assembling CPF records processes each EAD-encoded finding one at a time. While duplicate names (the name for the same entity occurring two or more times in the description) are merged before creating one CPF record for each unique entity found in the finding aid, duplicates that occur when the same named entity occurs in two or more finding aids are not. The process of matching and combining the duplicate CPF records is performed at SI/UCB.

Matching/Merging of CPF and VIAF Records – SI/UCB

The match and merge processing occurs in two steps. First, the CPF records are matched against one another, with records programmatically determined to be for the same entity combined. When combined, unique data in the two records are merged into the same record. Merged records may have two or more biographies or histories, will have two or more links to finding aids (integrated access), and will have an accumulation of all unique links to other named entities (social-professional-intellectual network). Second, the records resulting from the first step are matched against records in the VIAF. When records match, information from the VIAF record is integrated into the CPF record: alternative names, sex, associated country, language or languages used by the entity, and titles in WorldCat by the entity. A link to the VIAF record (same as) is also created, as well as a link to DBpedia when articles are available.

The most difficult challenge in this step in the processing is "identity resolution": one person can have more than one name and two or more people can have the same name. Names are not strong identifiers, particularly when the scope of the domain expands.

The technological foundation for the match/merge processing has two major components: Cheshire II (developed by Ray Larson), a high-performance XML-indexing platform that supports an array of advanced search algorithms; and MongoDB, an open source non-SQL database. The match/merge processing is completed in three phases: Load, Match, and Merge. In the Load phase, the system ingests

EAC-CPF records and records intermediary data into a MongoDB instance, including a unique record ID, name, existence dates, and other relevant pieces of context from the EAC record. This stage, in essence, creates a memory-backed cache for further query and merging operations. In the Match phase, the system conducts a multi-staged matching operation, applying different algorithms in each stage to determined matches. The benefit of this system is that additional stages can be inserted at any point, with the most accurate methods executed first. In this fashion, a matching operation cascades down the stages, being returned if a match is found at any point. The more (computationally) expensive matching operations do not execute until previous, presumably less expensive, stages have had a chance to make the match. In the final Merge phase, records containing data belonging to matched entities are merged together to form the final set of CPF records, for use in the development of the prototype public system developed by CDL.

The following provides a quantitative summary of the match/merge processing.

	Person	***Corporate***	***Family***	
Before Merge	125,555	47,189	2,899	175,643
After Merge	95,624	31,287	1,980	128,891

See appendix for a detailed description of the match/merge processing. (Larson, R.R., Janakiraman, K.: "Connecting archival collections: the social networks and archival context project. "In: *Proceedings of the 15th international conference on Theory and practice of digital libraries: research and advanced technology for digital libraries*. TPDL'11, Berlin, Heidelberg, Springer-Verlag (2011) 3{14} and Larson, R.R., and Liu, Yiming: "By Any Other Name: Name-based Record Disambiguation in the Social Networks and Archival Context Project' [unpublished paper])

Prototype Access and Historical Resource – CDL

Using the CPF records produced in the first two steps of processing, CDL has developed a prototype public archival access and historical resource. The prototype is based on the EXtensible Text Framework, and open source XML publishing platform developed at CDL. The prototype was different with five user personae in mind: a graduate student working on a PhD that involves biographies and the study of diplomatic families and networks; a person working at a repository that contributes records to the project who is interested in the benefits of the project for users; a library school student working on quality evaluation of record matching; archivist performing authority work while processing; and finally, an information technologist interested in Linked Open Data.

The prototype has two primary objectives. First, to provide integrated access to dispersed resources by and about a particular person, organization, or family. This access can be to known entities (for example, Richard Feynman), but also retrieve entities based on occupations and activities (Physicists) and the subjects of resources associated with entities (Quantum theory). The CPF record links to finding aids and bibliographic records that provide, in turn, information on the resources. Second, the biographical and historical information in each CPF, including links to related persons, organizations, and families, provide

uses with information on the life of the described entity, as the social-professional-intellectual network within which the entity lived and worked.

The following list provides a summary of the features of the prototype: 1) key word searching with completion assistance; 2) browsing ordered lists of names; 3) faceted qualifying of search results (occupation, subject, entity type, and so on); 3) biographies or historical information; 4) links to finding aids (entity creator of or referenced in described resources) and bibliographic records (for works by the entity); 5) links to associated persons, organizations, and families; 6) graphical (radial) display of the social graph; 6) links to VIAF records and DBPedia articles; and 7) exposure a subset of each CPF record as Linked Open Data.

CDL also produced an innovative "context widget" that users can install. With the widget installed in the users browser, the user will see an icon appear in the lower right-hand corner of the browser with the word "Context." Clicking on this icon will produce a list of the names appear in the finding aid that are in the SNAC prototype, with each name linked to the CPF record in the prototype.

For an overview of the features of the prototype and description of the technology used to develop the radial graph, see the appendix for "Historic Social Networks: Prototype Access System," a presentation at the annual conference of the Society of American Archivists by Brian Tingle.

Audiences

While the ultimate objective of the SNAC project is to dramatically improve access to and understanding of primary historical resources for researchers, teachers, students, and the general public, its immediate objective is to transform archival descriptive methods in order to facilitate meeting that objective. Thus the primary audience is the professional archival community, including manuscript librarians. Secondary audiences are library and museum professionals; archives, museums, and libraries, including consortia focused on providing access to cultural heritage resources; national and state cultural heritage institutions, in particular the Library of Congress, the National Archives and Records Administration (NARA), Smithsonian Institution, and state government archives; international standards organizations; resource description and access researchers; and service providers, OCLC in particular.

In order to reach this audience, the project collaborators have given public presentations in the U.S., Europe, and Australia (see the appendix for a complete list). The best evidence for demonstrating the success of the project in reaching the intended audience is the professional community's response to requests for contributing data to the Mellon-funded second phase of SNAC. In addition to making the VIAF available, OCLC also contributed 2.2 million MARC archival descriptions. The British Library, Bibliothèque nationale de France, Archives nationales (France), National Archives and Records Administration, the Smithsonian Institution have all contributed data. Thirteen consortia, twelve in the U.S. and one in the U.K. have contributed finding aids. Finally, 40 major research archives in the U.S. contributed finding aids. In all, over 150,000 finding aids and 300,000 archival authority records have been contributed to the SNAC project. (See http://socialarchive.iath.virginia.edu/contributors.html for a complete list.) Additional offers to contribute data continue.

Evaluation

SNAC was not formally evaluated by a third party. The lack of formal evaluation in the project work plan was a significant omission in the project work plan. The research work would have benefited in particular from three types of evaluation. First, a systematic review of the thoroughness of the extraction processing would have provided assurance that the techniques employed were correctly identifying and extracting relevant data. Second, a systematic review of the accuracy of the matching, scoring the processing with respect to true-positives, true-negatives, false-positives, and false-negatives, and further examining the likely causes of false-positives and false-negatives, would have provided very useful input for code and algorithm development. Finally, the development of the public prototype system would have benefited from formal user studies that engaged primary professional and scholarly audiences. Unfortunately, the budget was insufficient for accommodating these forms of evaluation, though such evaluation is now part of the second Mellon-funded phase of SNAC.

The software and processing results were evaluated on an ongoing basis by project staff. Though less than optimum, this technique produced useful confirming and corrective information over the course of the grant period.

The project, overall, has been very successful, meeting or exceeding expectations. The extraction, assembling, matching, and merging techniques have produced results of an acceptable quality even if not exceptionally high quality across the board. As one would expect, there is always room to improve algorithms for the key processes, but many of the qualitative issues can only be addressed by human editorial work. Identity resolution is problematic for computers and for people. The objective in developing systems such as SNAC needs to be pushing the boundaries of what computers can reliably perform, and where not, to develop systems that can assist human editors in efficiently and accurately solving the remaining problems.

The project failed to achieve several objectives. The software for extracting, assembling, matching, and merging was not released as open source as planned. The reason for this, as noted above, is that the developers felt that the code was insufficiently mature to share. More specifically, while it was functional, it was still very much under development. Both the extraction/assembling code, and the matching/merging code are undergoing thorough revision in the second phase of SNAC, and will be released as open source. The prototype public system code has been released as open source.

The plan of work also included developing Name Entity Recognition (NER) techniques for identifying names in archival description that were not tagged as names. While NER has been reasonably successful in natural language contexts, archival description (with a couple of exceptions) is not natural language. It might best be described as "semi-domesticated," which is to say, it is not sufficiently normalized to enable simple extraction techniques or natural, which would enable use of Natural Language Processing technique. While the project experimented with NER techniques, the results were not of sufficient quality to incorporate the code into the extraction/assembling processing.

Continuation of the Project

During the grant period, IATH successfully applied for two grants, one to continue the SNAC research and development work, and the other to develop the authorities description expertise of the archival community *and* to develop a blueprint for a national archival authorities cooperative.

The Andrew W. Mellon Foundation has funded SNAC to expand the research agenda, increase vastly the quantity and diversity of the source data, and develop a data and technological foundation on which to establish a national archival description and access cooperative. In this second phase, the EAD-encoded finding aids have been expanded from 30,000 to 150,000 or more. These finding aids are made available to the project by thirteen consortia and nearly forty research archives. In addition, OCLC WorldCat has made available 2.2 million MARC archival descriptions. NARA, British Library, New York State Archive, Smithsonian Institution, BnF, and Archives nationales (France) have contributed over 300,000 authority records. Given the diversity of the source data, extraction and conversation software is being developed to assemble CPF records not only from EAD-encoded finding aids, but also from MARC archival descriptions and a variety of SQL databases that are used for natively maintaining authorities data. Many of the processes in the first phase of SNAC are also being rewritten to support incremental processing of incoming CPF records, and to provide a technological foundation for a description cooperative maintenance system authorities data. User studies are also being used to provide input for a thorough revision of the public prototype.

In July of 2011, IATH, in collaboration with Simmons College, applied for funding from the Institute of Museum and Library Services for "Building a National Archival Authorities Infrastructure." This project has two objectives. First, it is offering 140 scholarships, to be used for regional Society of American Archivists workshops introducing the professional community to EAC-CPF. Katherine Wisser (Simmons College) is managing the dispersal of the scholarships and is teaching the workshops. Second, the project is developing a blueprint for an archival access cooperative, similar to the Program for Cooperative Cataloging that is managed by LoC. As currently conceived, the cooperative will initially focus on cooperative archival identity description (archival authority control). The National Archives and Records Administration (NARA) has formally agreed to host the cooperative and to manage its administration and governance. The technical infrastructure will be collaboratively developed and hosted outside of NARA. The planning for the cooperative has been greatly facilitated by Laura Campbell (retired CIO of LoC), Don Waters (Andrew W. Mellon Foundation), Clifford Lynch (CNI), Anne Van Camp (Smithsonian Institution Archives), John Martinez (NARA) and Jerry Simmons (NARA). Before the work on establishing the cooperative begins, the Mellon Foundation has invited a proposal for a one-year planning proposal (to be submitted later at the end of 2013) that will establish the charter within NARA, determine in detail the technical requirements, resolve intellectual property issues, and conduct user studies that will inform the technological development of the cooperative.

Long Term Impact

As described above, SNAC will lead to establishing a sustainable archival description cooperative. For NARA, this will provide a leadership role that the professional community has long desired. For the

professional community, it will offer an unprecedented opportunity to better and more efficiently describe its holdings, to provide union access to its holdings, and at the same time to build a historical resource that will provide in-depth access to the social-historical context within which the records were created. With respect to the impact on users of archival holdings, an observation made by Edward Ayers, President of the University of Richmond and a member of the SNAC Advisory Board, states it most succinctly: " SNAC promises to change the way history is imagined and written."

Please see the appendix for additional comments from Advisory Board members.

Grant Products

On-line resources

- http://socialarchive.iath.virginia.edu
- http://socialarchive.iath.virginia.edu/prototype.html
- http://socialarchive.iath.virginia.edu/xtf/search

For the following open source software web services see
http://socialarchive.iath.virginia.edu/software.html

1. Extracting data from MARC archival descriptions and assembling CPF records.
    a. Web service
    b. Software for local use
2. Prototype Access System
    a. Core site: XTF
    b. Graph building and rendering
    c. SNAC "Context" widget
3. SPARQL endpoint

The project plans to release additional software used in the creation of CPF records and all of the software used in the prototype access and resource.

The SNAC collaborators will continue to find professional venues within which to present project reports, and to write additional papers for publication.

Appendices

List of research and presentations

Related Research

Krishna Janakiraman and Sean Marimpietri based their Masters research on SNAC at the School of Information, UC Berkeley, spring semester 2011.

Thomas Lynch, graduate student at the Graduate School of Information and Library Science, University of Illinois, Urbana-Champaign, held an internship at IATH (spring and summer 2011) focused on identifying personal names in titles that are not tagged as personal names. The goal was to increase the number of names derived from correspondence description.

SNAC data was provided in August 2011 to the NEH-funded Research-oriented Social Environment project (RoSE, http://rose.english.ucsb.edu/). Project director Alan Liu, University of California, Santa Barbara.

Presentations and Publicity

2010

August 9: Pitti, at the EAC-CPF: Moving Forward with Authority workshop, National Archives and Records Administration, Washington, D.C.

October 1: Larson, Tingle, and Janakiraman, Friday Seminar, School of Information, UC Berkeley.

November 3: Tingle, Digital Library Federation (DLF) Forum 2010, Palo Alto, California.

November 4: Pitti, presentation to the National Archivist and the Advisory Committee on the Electronic Records Archive, National Archives and Records Administration, Washington, D.C.

2011

January 7: Larson, "Web Information Quality and Name Authority Control," Evaluating Web Information Quality and Credibility Panel, Hawaii International Conference on System Sciences.

February 4: Larson, Friday Seminar, School of Information, UC Berkeley.

February 25: Larson, Personal Archiving Workshop, San Francisco.

March 7: Pitti, Archivists Roundtable of Metropolitan New York.

May 25: Pitti, Larson, and Tingle, OCLC Research staff (Dublin, Ohio; San Mateo, California). Some slides available at http://www.slideshare.net/tinglebrian/snac-oclcmay2011

May 23: Turner and Tingle, Webinar, for archivists and librarians affiliated with the Online Archive of California. URL: http://www.cdlib.org/services/dsc/webinars/snac/lib/playback.html

May 12: Janakiraman, BioGraph, Masters in Information Management and System Final Project Presentation, School of Information, University of California, Berkeley.

June 22: Pitti, Janakiraman, Larson, and Tingle, "The Social Networks and Archival Context Project," Digital Humanities 2011 conference. URL: http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-261.xml;query=ray%20larson;brand=default

August: SNAC Project, SAA 2011 Description Expo entry (http://www2.archivists.org/groups/description-section/description-expo-2011)

August 27: Pitti, Larson, and Tingle (Michele R. Combs, Chair), "The Social Networks and Archival Context Project: EAC-CPF at Work," SAA annual meeting

September 26: Larson and Janakiraman, "Connecting Archival Collections: The Social Networks and Archival Context Project," Theory and Practice of Digital Libraries, Berlin (http://www.tpdl2011.org/images/stories/tpdl/TPDL_Larson.pdf)

2012

February 10: Larson and Tingle, "Update on the Social Networks and Archival Context (SNAC) Project," UC Berkeley School of Information, Friday Afternoon Seminar series

April 4: SNAC Presentation (Pitti) The Management of Scholarly Identity: A CNI Workshop (Coalition for Networked Information)

May 13: "Projects Aims to Build Online Hub for Archive Materials" by Jennifer Howard, *Chronicle of Higher Education* http://chronicle.com/article/Building-a-Digital-Map-of/131846/

May 21-23: SNAC presentation (Pitti, Larson, Tingle, Turner) Building a National Archival Authorities Infrastructure meeting, at NARA

June 21: “Introduction to Linked Data: Connecting Our Collections” (Tingle, Turner, Wisser) Rare Books and Manuscripts Section (Association of College & Research Libraries) Preconference, San Diego, CA

July 17 Digital Humanities Conference Hamburg, Germany "Facilitating Research through Social-Document Networks" panel organized by Daniel Pitti (with Simon Agnès, BnF; Stefano Vitali, Soprintendenza Archivistica per l’Emilia Romagna; Arnold, Kerstin, Bundesarchiv and APEx)

August 11: “Solving Our Problems with Authority and Sharing: Current Developments and Prospects” (Larson and Pitti) Society of American Archivists, San Diego, CA

August 22: "Social networks and archival context: findings and future research" (Pitti) ICA Congress, Brisbane, Australia

November 4: “Social Networks and Archival Context Project” (Larson, Tingle, Turner) Digital Library Federation, Denver

2013

February 21 (Paris), 25 (Strasbourg), and 27 (Berlin): Presentations on SNAC at Archives nationales (France); VIAF Council meeting; and APEx meeting (Pitti)

SNAC Advisory Board Statements

Ed Ayers
President of the University of Richmond
Professor of American History

SNAC promises to change the way history is imagined and written. For all that the digital revolution has revolutionized, the heart of research lies within the primary record embedded in archives large and small. The pioneering work of SNAC will unlock that record, revealing connections and patterns invisible to us now.

Sue Perdue
Director of People of the Founding Era
Virginia Foundation for the Humanities

One of the biggest challenges facing documentary editors, and historians more generally, is the task of locating manuscript collections for a given individual. More daunting still is searching for people who show up in manuscript collections as peripheral to the main person in the collection. This is the beauty of SNAC. It aggregates the incredibly valuable information locked in digitized finding aids related to these peripheral people, and pulls it into a single access point. The promise of a national effort to make this a standard for all archives is significant. A national archival authority will be a major boon to scholars because it will lead them to the known--and better still--the unknown archival collections.

Ken Price
Hillegass University Professor of American Literature
University of Nebraska

SNAC reshapes how we think about creative acts. Documents and inventions, too often seen as the product of an isolated mind, are almost always better understood as resulting from social networks. Take the case of Walt Whitman. He is sometimes referred to as *The Solitary Singer* (the title of a famous biography). But to focus on Whitman alone is to risk missing the key importance of the bohemian group that gathered at Pfaff's beer hall in New York and advanced Whitman's career. Later he profited from another under-studied group of government clerks, many of them fellow writers, who sustained him with their friendship and sometimes collaborated with him in Washington, DC. A project like SNAC moves us away from the myth of the solitary genius and provides tools for seeing and studying the social network out of which people work and create.

Alan Liu
Professor of English
University of California, Santa Barbara
Director of Research oriented Social Environment (RoSE)

SNAC is a premier example of using digital methods to create a next-generation knowledge resource for historically-deep and internationally-broad research. SNAC employs state-of-the-art computational techniques to do three things very well:

- unlock information originally recorded for specific purposes in library and other archival finding aids to make them usable in new contexts;
- connect widely-distributed information of this sort from around the world;
- and marry the "library" or "archive" model of knowledge to a whole other model of social networks that both humanizes our understanding of the way knowledge emerges from communities of knowledge creators and seekers and speaks powerfully to today's "social network" generation.

The resulting whole is much more than the sum of the parts. To build on the extractive metaphors such as "data mining" or "digging into data" often used for the new information technologies, SNAC could be described as knowledge "fracking": a next-generation means of releasing unrealized knowledge from individual strata so that they can pool together and vastly multiply the value of the original knowledge deposits.

For the individual researcher and student, SNAC will be an invaluable means of tapping into far-flung knowledge in ways that exceed the capabilities of existing archival exploration systems. For future institutional and other projects, SNAC will serve as an upstream source of knowledge for their downstream client knowledge systems. For nations and states, SNAC will join such important projects as the "Digital Public Library of America," "Your Paintings" in the United Kingdom, "Europeana" in the European Union, and "HUNI" in Australia in assembling and making accessible the aggregate pool of their people's cultural heritage.

Michael Rush
Processing Archivist, Beinecke Rare Book and Manuscript Library
Yale University
Co-chair, Technical Subcommittee-Encoded Archival Description (Society of American Archivists)

From the perspective of a practicing archivist, a cooperative to assemble descriptions of people and groups with links to the archival records that document their activities will be a boon to the efforts of archivists everywhere. It will connect my collections to related ones elsewhere, and it will help connect end-users with collections everywhere. I anticipate that a central resource for the description of the people and groups documented in archives will result in a measurable economy of scale, promoting less duplication of effort both by archivists engaged in description and users engaged in research.

Jerry Simmons
Specialist for Data Standards / Authority Team Lead
U.S. National Archives and Records Administration

SNAC provides multi-path entry for discovering knowledge in archives by exploring identity descriptions of creators of primary source materials, whether a person, a family, or a corporate entity. It offers the greater community of researchers free and easy access to a great and growing store of archival creator information. A sure benefactor is the library/archives name authority cataloger. No matter their affiliation, name catalogers can use SNAC to discover biographies and administrative histories to support their name

cataloging efforts.

Then creator name discovery leads to linking to other identities. NARA name authority catalogers look forward to contributing identity descriptions for creators which can be linked to other persons and to federal agency names, illustrating the every important contextual relationship between persons and federal agencies and the agency's records. Ongoing efforts to establish a global archival authority cooperative will only increase benefits for archivists and catalogers as the network of contextual information about archives and archives creators continues to grow through crowd cooperation.

Michael Fox
Director of the Minnesota Historical Society (retired)

As the Social Networks and Archival Context project works toward the creation of a National Archival Authority service, the multiple benefits of this initiative are already apparent. A substantial body of biographical and historical information, presently dispersed and often difficult to locate, about the individuals, organizations and families whose activities are documented in nation's historic record is coming together in a single repository of information.

The benefits for research of this initial step are readily apparent. But as useful as this repository will be in its own right, it has the potential for far greater impact, both for archivists in their daily work and for researchers, serving first as a portal to the archival finding aids that provide fuller documentation of the papers and records of these individuals and organizations, and then perhaps ultimately as the gateway to digital copies of the documents themselves, now often dispersed across multiple repositories.

Anne Van Camp
Director of the Smithsonian Institution Archives

The formation of a National Archival Authority Cooperative is an important effort to aggregate data about people, families, and organizations that is often hidden in traditional archival descriptions. Repositories that serve as stewards of archival material rarely have the resources to adequately describe the creators of the records they hold. By harvesting and aggregating this data from a variety of sources, the database will become a real boon to researchers and to archivists as well. Researchers will be aided in finding new sources of information that were previously inaccessible, and archivists will be able to add to and borrow from an authority file that will begin to standardize names for archival creators.

Connecting Archival Collections: The Social Networks and Archival Context Project

Ray R. Larson and Krishna Janakiraman

School of Information
University of California, Berkeley
Berkeley, California, USA, 94720-4600
ray@ischool.berkeley.edu, krisha@ischool.berkeley.edu

Abstract. This paper describes the Social Networks and Archival Context project, built on a database of merged Encoded Archival Context - Corporate Bodies, Persons, and Families (EAC-CPF) records derived from Encoded Archival Description (EAD) records held by the Library of Congress, the California Digital Library, the Northwest Digital Archives, and Virginia Heritage, combined with information from name authority files from the Library of Congress (Library of Congress Name Authority File), OCLC Research (The Virtual International Authority File), and the Getty Vocabulary Program (Union List of Artist Names). The database merges information from each instance of an individual name found in the EAD resources, along with variant names, biographical notes and their topical descriptions. The SNAC prototype interface makes this information searchable and browseable while retaining links to the various data sources.

1 Introduction

One of the important tasks of scholars is to use secondary (e.g. books and journal articles) and primary (original manuscripts, letters, etc.) research information in examining the lives, work, and events surrounding historic persons. For historians and many other scholars, the prefered sources are primary – the actual words or documents of the persons or organizations concerned, and to a lesser extent the interpretations of other scholars about those persons or organizations. In digital library research the focus has been largely on these secondary resources, and not on the original primary resources. In part this has been due to the relative lack of available metadata and digitized content for primary resources when compared to those of secondary resources.

The Social Networks and Archival Context (SNAC) project is trying to address this challenge of improving access to primary humanities resources through the use of advanced technologies. The project is producing software and developing open linked data resources that will enable scholars to connect historic persons to existing archival descriptions and to library catalogs and authority files. This is creating a powerful new resource that enhances access to and understanding of the cultural resources in our archives, libraries, and museums

through the description of the people who created them and whose lives are reflected in those resources. This new resource can serve a wide variety of objectives to benefit scholars, educators, students, and anyone interested in the record of our past. The goals of the SNAC project are to:

1. Support scholars and other users in discovering and identifying persons, families, and organizations, by making the names used by and for them searchable.
2. To merge together information from a wide variety of different sources by and about people and organizations and thus to enhance access to primary and secondary resources.
3. To discover and provide access to the social and professional networks within which people lived and worked by systematically documenting their relationships with one another and making this documentation available to users in order for them to better understand the social-historical contexts within which resources were created.
4. Provide archives, libraries and scholars with access to records describing persons, families, and organizations, thereby improving description of archival records and creating efficiencies in the re-use of metadata across repositories, and through open linked data resources (Figures 1, 2 and 3 show screens from the our prototype interface).
5. Connect traditional library and archival information on persons, families and organization with semantic web resources on the same persons, families and organizations, providing a resource for validation and contextual matching of traditional and semantic web resources.
6. Make available the software developed for matching persons, families and organizations based on the approaches developed during the project for name matching that takes into account the contextual information associated with the name to provide more effective determination of identity between different sources for information on individuals, families and organizations.

At the core of the SNAC project are the Encoded Archival Context - Corporate bodies, Persons, and Families (EAC-CPF) and the Encoded Archival Description (EAD) XML markup standards. SNAC is developing open-source software that will facilitate efficiently and accurately deriving authority control records from existing EAD archival finding aids from the Library of Congress (LoC) and three consortia, the Online Archive of California (OAC), the Northwest Digital Archive (NWDA), and Virginia Heritage (VH), and enhancing them with additional information in matching LoC, Getty Vocabulary Program, and Virtual International Authority File (VIAF) name authority records.

The SNAC project is being led by the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia, in collaboration with the California Digital Library (CDL), and the University of California, Berkeley School of Information (SI). The SNAC prototype access system being developed by the CDL is now publicly available to demonstrate how the descriptions of persons, families, and organizations - a traditional and integral part of archival

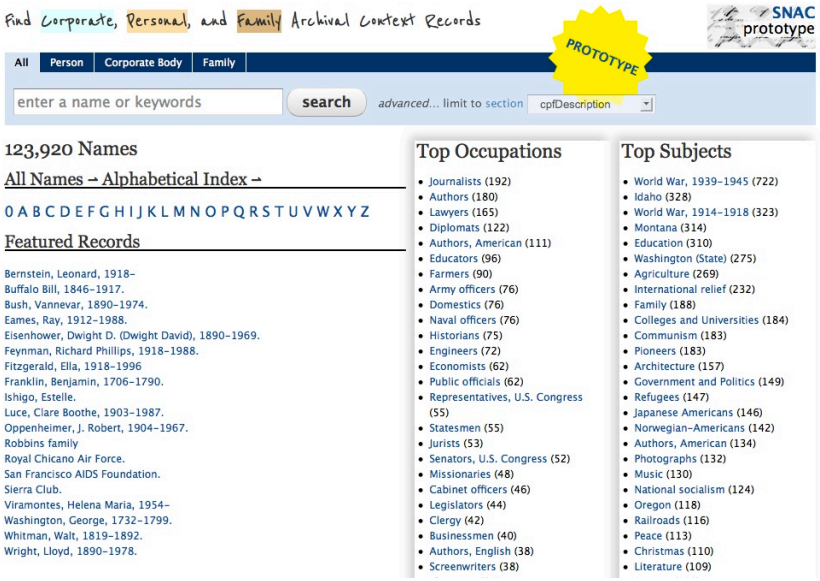


Fig. 1. Browsing Screen for SNAC Prototype

practice can be used to facilitate efficient access to and understanding of not only archival resources, but also library and museum resources. The SNAC project is intended to benefit the humanities community most broadly, as it will assist the work of both archivists and users.

In the remainder of this paper we will describe the processes involved creating the SNAC database and how it is presented through the public interface. The next section describes the extraction of EAC-CPF records from EAD records. We then examine how the names of individuals, organizations and families are matched across the different sources of EAD data, and how they are matched with the library authority records from the Library of Congress and the Virtual International Authority File. In addition we present the results and effectiveness of different matching methods. Finally we discuss how the records are indexed and presented in the prototype public interface.

2 Extracting EAC from EAD

As mentioned above, the basic EAC-CPF records used in the SNAC project are derived from EAD archival descriptions. The current approach to doing this extraction uses the Extensible Markup-Language Transformation (XSLT 2.0) language in conjunction with the XPath 2.0 standard. Through the use of regular expressions and specially designed functions, the XSLT transform identifies elements of the EAD that represent individual persons, corporate bodies and families in various parts of the EAD record. Currently we have focused on the identification and extraction of individual records from the following EAD tag

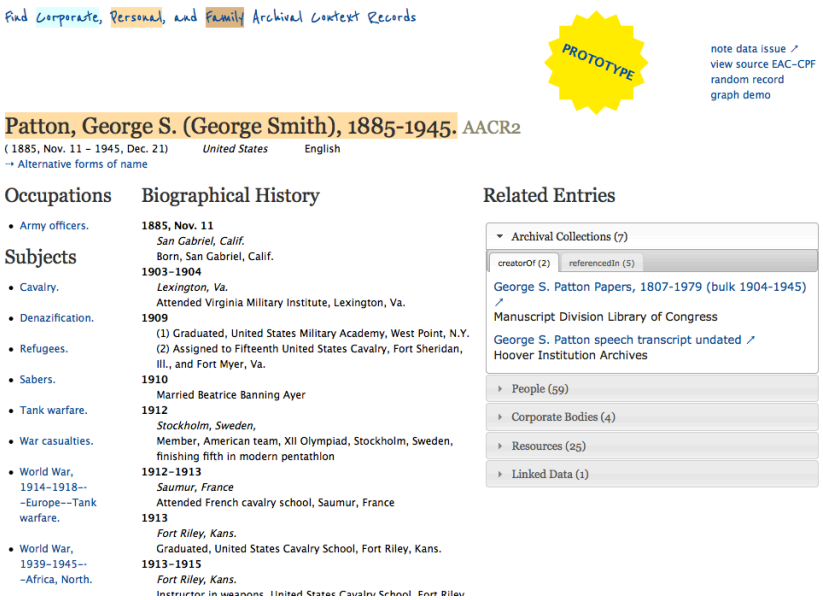


Fig. 2. Example Record (George S. Patton)

components: <persname>, <corpname>, and <famname> that occur within <origination>, <controlaccess>, and <unittitle>.

In the EAD records that we are using, the contributing archives have followed the "best practices" for encoding the <origination> and <controlaccess> elements, so these are usually formulated following strict cataloging rules (AACR2, for American archives and libraries). Most of our difficulties (in matching names and extracting contents) are caused by names that have not been formulated according to such rules. In some cases, for example, names are presented in direct order (John Smith) rather than inverted order (Smith, John), and may also be combined data that is not part of the name, (including subject subdivisions or uniform titles).

For each unique name string extracted by the above process, an EAC-CPF record is created. For records derived for creators (i.e., the source of the archival records), additional descriptive data for dates of existence, occupation, subject headings assigned to records, languages used, and biographical-historical information is extracted into the corresponding EAC-CPF records.

Once the EAD records have been processed, the result is a set of EAC-CPF records each containing a single identified name along with identification of the source EAD, and, in the case of creator records, any biographical information, dates of existence, etc. from the EAD source. The EAC-CPF records can encode extensive information about an entity, drawn from various parts of the source records. In addition to basic identifying information (name, type, occupation(s), and existence dates), they include an entity's relationship(s) with other entities,

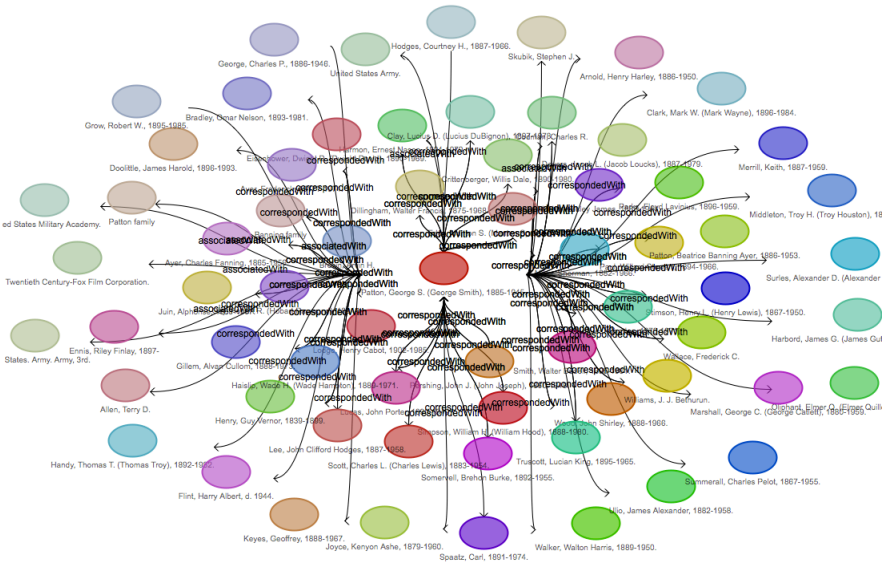


Fig. 3. Social Network (George S. Patton)

resources, and works. Since EAC-CPF records are derived independently from each EAD record, there can be multiple records representing the same entity in multiple EAD collections. A key problem, then, is to identify multiple EAC-CPF records that represent the same entity and merge them together into a single record.

3 Matching Names in EAC-CPF and Authority Files

The next stage in processing the extracted EAC-CPF records attempts to combine the information of the individual records for the same entity into single records, and to enhance them with information derived from library authority files. This process involves matching the names in the EAC-CPF records with each other, and with the records in library authority files.

We currently have 158,079 EAC-CPF records114,639 persons, 41,177 corporate bodies and 2263 family namesderived from Library of Congress, the Online Archive of California and North West Digital Archive EAD records. The records were parsed with the EAC-CPF specification to extract information on entitys name, type and relations, stored in a relational database. Preferred and alternate names from the VIAF name authority files were indexed using the Cheshire II information retrieval engine [3], which uses a probabilistic information retrieval algorithm to find the top n VIAF records and their associated names given an entity name. We mapped each EAC-CPF entity to names from top n VIAF records (currently the top five VIAF records). These mappings are also represented in a relational database.

Our problem is similar to the well-studied named entity disambiguation problem, where the task is to identify the correct entity, in a given context, from a set of similar entities. Standard approaches use statistical learning techniques, either performing supervised learning and train classifiers that predict the relevance of an entity given a context or performing unsupervised learning and design clustering techniques that cluster similar entities together. As an example of the former, Bunescu and Pasca [2] suggest a method that trains Support Vector Machines (SVM) classifiers to disambiguate entities using the Wikipedia corpus. The classifier was trained using features extracted from the title, hyperlinks linking other entities, categories assigned to the entity and Wikipedias redirect and disambiguate pages. Bagga and Baldwin [1], and Mann and Yarowsky [4], are examples of the latter technique, where similar entities are clustered using features extracted from entitys biographical information, words from sentences surrounding the entity in texts and entitys social network and relationships. Other techniques involve using gazetteers and name authority files as external references to aid the disambiguation process. Smith and Crane [5], for example, use gazetteers to disambiguate geographic place names. In the remainder of this section we describe some experiments in name matching using different approaches.

3.1 Experiments in Supervised Name Disambiguation

Given a collection of entity/name pairs and a query name, the supervised name disambiguation problem is to predict the entity that the query name is associated with. We describe a supervised learning approach, based on the Naive Bayes classifier as an approach to solving this problem. Given an exhaustive collection of entities with a set of possible names for each of these entities, we compute simple representation of these names using shingles (also called character ngrams) and train naive bayes classifiers that predict the entity a name belongs to using these representations.

We represent each name as a list of l length shingles. The shingles are computed by using a sliding window that is l characters in length. For example, the $l = 3$ shingles for the name Albert Einstein would be alb, lbe, ber, ert rt_ ,t_e, ... ein (where "_" represents white space between words).

This representation allows us to consider a name as a vector in a k dimensional vector space defined by the shingles, where k is the number of l length shingles in our collection. In addition to abstracting the structure of a name, this representation also allows us to create a shingle - entity reverse index. As would be explained in later sections, this index allows us to significantly reduce the space of candidate entities when a query is made.

3.2 Approaches

String Edit Distance. Our first approach uses edit distance to rank entities based on a query name. Given a query name, we rank entities based on the mean edit distance between the names associated with the entity and the query name.

We used the Levenshtein Distance to compute the edit distance between two strings.

Computing the edit distance over the entire space of entities is not feasible. To reduce the space, we only consider those entities that are indexed by the shingles present in the query name. Specifically, we first rank our entities based on how many shingles in the query name are indexed by them. We then pick top-M of these entities and score them based on the mean edit distance between the names associated with the entity and the query name. Entities with lower mean edit distance are ranked higher.

Naive Bayes. If we consider each entity as a class and names associated with the entity as examples of the class, we can cast the name disambiguation problem as a supervised statistical learning problem. Specifically, given a query name x_q, our problem then is to find an entity class C that is mostly likely to have generated the query name.

$$C = \arg\max_i P(C = c_i | X = x_q) \tag{1}$$

Using the Bayes rule, we can estimate the most likely C as follows,

$$C = \arg\max_i \frac{P(C = c_i)P(X = x_q | C = c_i)}{P(X = x_{q)}} \tag{2}$$

As $P(X = x_q)$ will be the same for all the entity classes, we can eliminate it from the above equation,

$$C \leftarrow \arg\max_i P(C = c_i)P(X = x_q | C = c_i) \tag{3}$$

The prior probability $P(C = c_i)$ can be estimated as,

$$P(C = c_i) = \frac{\#\{C = c_i\}}{D} \tag{4}$$

where D is the total number of name instances in the training collection and $\#\{C = c_i\}$ represents the number of name instances with entity class as c_i.

Based on our representation of names using shingles, we can consider each name as an instance of random vector in a k-dimensional space of shingles. The posterior probability in the above equation can then be written as,

$$P(X = x_q | C = c_i) = P(X = (x_{q1}, x_{q2}, x_{q3,}..., x_{qk}) | C = c_i) \tag{5}$$

where x_{qj} represents the event that a particular shingle j occurs in the query name. Estimating the above posterior probability would amount to estimating the joint distribution of the random vector X , assuming each component of the k -dimensional random vector to be binary, this would amount to learning 2^{k+1} parameters.

The Naive Bayes classifier reduces the number of parameters to learn by assuming conditional independence, that is components of the random vector X

are conditionally independent given C . Under this assumption, the posterior probability can be computed as,

$$P(X = x_q|C = c_i) = \prod_{j=1}^{k} P(X = x_{qj}|C = C_i) \tag{6}$$

Once the posterior probabilities are estimated, the most probable entity for a given query name X_q can be estimated as,

$$C \leftarrow \arg\max_i \ log(P(C = c_i)) + \sum_{j=1}^{k} log(P(X = x_{qj}|C = c_i)) \tag{7}$$

We used two different approaches for computing the posterior $P(X = x_{qj}|C = C_i)$.

Shingle Presence. Our first approach considers X as a binary random vector. Specifically, if X_n is an outcome of the random vector that represents a name n, then $X_{nj} \in \{0, 1\}$ depending on whether the shingle corresponding to the j^{th} component is not present or present in the name. Given this formulation, the posterior $P(X = x_{ni}|C = C_i)$ can be estimated by computing the frequencies of a shingle's presence for the given entity class,

$$P(X = x_{nj}|C = c_i) = \frac{\#\{x_j = 1 \cap C = c_i\} + \delta}{\#\{C = c_i\} + k\delta} \tag{8}$$

where δ is the Laplace smoothing factor and k is the total number of shingles derived from the training set.

Multinomial Model. For our second approach, instead of restricting each X_{nj} to be binary, we let X_{nj} to represent the number of times the shingle corresponding to the j^{th} component occurs in the given name. The posterior $P(X = x_n|C = C_i)$ can then be considered as the multinomial distribution, with the shingles as the different categories. The probability of occurence of a shingle given a entity class can then be estimated as,

$$P(X = x_{qj}|C = c_i) = \frac{\#\{x_j \cap C = c_i\} + \delta}{\sum_{j=1}^{k} \#\{x_i \cap C = c_i\} + k\delta} \tag{9}$$

where δ is the Laplace smoothing factor and k is the total number of shingles derived from the training set

3.3 Experiments

Dataset. We used the VIAF collection for testing our approaches. The VIAF collection consists of name authority records for around 2.5 million entities. Each

entity record in the VIAF collection consists of a set of names that could be associated to the entity and may also include other information such as gender, nationality, and birth and death dates, if available. We filtered the VIAF dataset by first removing names that use non-ascii characters above a defined threshold. This eliminated most names in languages other than those in Western European Languages. We further filtered this dataset by removing entities with less than 4 names. These filtration steps reduced our dataset to 291,952 entities and 1,886,049 names. Names were normalized by transforming to lower case and further by removing all punctuation except space. We retained space as shingles containing the space naturally modeled the components of the name. We also removed existence dates from the names as this information was available separately in the VIAF record of the entity corresponding to the name.

Experiment design. To test our classifiers, we split our collection of name instances from the VIAF dataset into training and testing sets using a 70-30% split ratio. This gives 291,952 entity classes with 1,320,234 training instances and 565,815 testing instances. Each training and testing instance includes a name, the entity class it belongs to, and birth and death dates. We used a shingle length of 3 characters to compute our feature representations, for our training set this amounted to 14,103 unique shingles.

Given a query name from the test set, we predict its entity class and compare it with the actual entity class the name belongs to. To compare our classifiers, we used the %accuracy measure. In addition to considering a prediction to be accurate if the correct entity is ranked first, we also considered the cases when the correct entity comes within top-5 or top-10 of the ranked list.

We use a reduced search set by ranking entities based on how many shingles in the query name are indexed by them and considering only the top-M entities of the ranked list. For all our results below we used M = 20.

Results. Our first objective was to predict the entity a query test name belongs to using only the strings present in the name. Table 3 shows the accuracy results for our three approaches. The Edit distance based approach is attractive as no model estimation, other than builing the shingle reverse index, is neccessary. However, they did not perform as well as the probabilistic approaches. The main reason was that edit distance is sensitive to the order in which the components of the name occur. For example, a sample of different names for the physicist Richard Feynman are shown in Table 1. A majority of the names start with the lastname Feynman followed by the firstname Richard. If the query name is 'Richard Feynman' the average edit distance between the query name and the names in the collection would be rather high than compared to a name such as 'Richard Gere'. The probabilistic approaches handle this issue as the shingle features capture information that is much more local than the entire string of the name. The multinomial model works better than the shingle presence model as it captures the frequency of a shingle's occurence, rather than its mere presence, in the names associated with an entity. Table 2 shows the top-5 results for the

Table 1. Alternate names for the Physicist Richard Feynman

Alternate names for the Physicist Richard Feynman
Feynman, Richard P
Phillips Feynman, Richard
Feinman, Richard P
Feynman, Richard
Feynman, Richard Phillips
Feynman, R. P.
Feynman

query name 'Richard Feynman' for the three approaches. The names shown were randomly sampled from the entity classes. Although the Multinomial Model

Table 2. Top-5 results for the query name 'Richard Feynman'

Edit Distance	**Shingle Presence**	**Multinomial Model**
Norman, R. J	Calichman Richard F	Harman, Richard Michael...
Staar, Richard Felix	Harman, Richard Michael...	Calichman Richard F
Calichman Richard F	Babcock, Richard Felt	Feynman, Richard P
Manby, Richard,	Kern, Richard	Manby, Richard
Kern, Richard	Kahn, Richard	Kahn, Richard

Table 3. Accuracy results for automatic name disambiguation using only the name strings

Approach	**Edit Distance**			**Shingle Presence**			**Multinomial Model**		
Criteria	*First*	*top-5*	*top-10*	*First*	*top-5*	*top-10*	*First*	*top-5*	*top-10*
%Accuracy	42.9%	72.18%	83.4%	43.12%	74.87%	83.3%	60.82%	82.49%	86.71%

is able to predict the correct entity with reasonable accuracy, for automatic name matching we would need the top-ranked entity to be correct with a much higher accuracy. This motivated us to use the existence dates as an additional information. To do this, we boosted the scores given to the entity classes by our classifiers if the query name's existence dates matched with the existence dates associated with the entity class. We used the following rules to boost the classifier scores.

$$score = \begin{cases} score + \gamma & \text{both dates match} \\ score + \gamma/2 & \text{either birth or death dates match} \\ score & \text{dates do not match} \end{cases}$$

Instead of checking if the dates exactly matched, we assumed two dates to match if the differance between the dates was less than 5 years. This, we felt, would handle cases when there were reasonable variations in the existence dates. Table 4 shows the top-5 results for the query name 'Richard Feynman' for the three

approaches with the scores boosted using existence dates information. The names shown were randomly sampled from the entity classes. Table 5 shows that the

Table 4. Top-5 results for the query name 'Richard Feynman' using existence dates information

Edit Distance	Shingle Presence	Multinomial Model
Feynman, Richard P	Kahn, Richard F	Feynman, Richard P
Kahn, Richard F	Feynman, Richard P	Kahn, Richard F
Hamann-MacLean, Richard	Hamann-MacLean, Richard	Hamann-MacLean, Richard
Staar, Richard Felix	Babcock, Richard Felt	Babcock, Richard Felt
Babcock, Richard Felt	Staar, Richard Felix	Staar, Richard Felix

accuracy of our classifiers significantly improved. These results were obtained using $\gamma = 100$

Table 5. Accuracy results for automatic name disambiguation using name strings and existence dates

Approach	**Edit Distance**			**Shingle Presence**			**Multinomial Model**		
Criteria	*First*	*top-5*	*top-10*	*First*	*top-5*	*top-10*	*First*	*top-5*	*top-10*
%Accuracy	80.8%	89.14%	89.47%	84.72%	89.47%	89.49%	80.21%	89.25%	89.49%

4 The Prototype SNAC Interface

The prototype public interface for the SNAC EAC-CPF database was developed by the California Digital Library and uses their open source eXtensible Text Framework (XTF) system to support search, display and navigation of the EAC-CPF records.

As shown in Figure 1 the user is able to browse the people, organizations, and families in the SNAC collection alphabetically by name, by occupation (for persons) or by subjects associated with the source EAD collection. Some interesting example records are also made available as "Featured Records" on this initial screen, which also provides a search capability that can access any of the content of the records (with names weighted more highly in the results).

Figure 2 shows a single record for the American General George S. Patton. The prefered form of the name for entries is shown (Alternate forms of the name from library authority files are available as a pop-up window). The record also shows (on the left-hand side) the occupations, and topical subject headings associated with the entity. If one or more of the source EAD records includes a biographical entry it is included also. On the right-hand side of the page access is provided to the archival collections created by the named entity (creatorOf), or in which the entity is referenced (referencedIn). The same panel provides access to other people and corporate bodies associated with the named entity. These are often the correspondents, family members or others referenced in the EAD

sources. This data is also the basis for the constructing the social network of the named entity as shown in Figure 3 (this graphical depiction of the social network is accessed through the "graph demo" link at the top right). Resources such as books by or about the named entity are also included as are linked data resources, such as links to the library authority record for the named entity.

5 Conclusions

This paper has described the Social Networks and Archival Context project, and examined some of the issues and processes in deriving an authority database for archival collections from EAD records using the EAC-CPF format. We examined some of the issues in matching and merging named entities from different collections and reported on some of our experimental work in name matching. Finally we described the prototype interface for public use of the SNAC database generated by the processing, matching and merging of named entities.

The SNAC project is still in progress, and we are now starting to apply the results of our research (and our ongoing analysis of matching and merging failures) to improved versions of the database.

6 Acknowledgments

The work presented in this paper is based on the Social Networks and Archival Context project (Daniel Pitti, Principal Investigator) funded by the (U.S.) National Endowment for the Humanities at the University of Virginia Institute for Advanced Technology in the Humanities (IATH), the California Digital Library, and the University of California, Berkeley School of Information. Information on the SNAC project, and access to the prototype is available at http://socialarchive.iath.virginia.edu/ .

References

1. A. Bagga and B. Baldwin. Entity-based cross-document coreferencing using the vector space model. In *Proceedings of the 17th International Conference on Computational Linguistics, Volume 1*, pages 79–85, 1998.
2. R. Bunescu and M. Pasca. Using encyclopedic knowledge for named entity disambiguation. In *Proceedings of EACL*, page 6, 2006.
3. R. R. Larson, J. McDonough, P. O'Leary, L. Kuntz, and R. Moon. Cheshire II: Designing a next-generation online catalog. *Journal of the American Society for Information Science*, 47(7):555–567, July 1996.
4. G. S. Mann and D. Yarowsky. Unsupervised personal name disambiguation. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003-Volume 4*, pages 33–40, 2003.
5. D. A. Smith and G. Crane. Disambiguating geographic names in a historical digital library. In *Research and Advanced Technology for Digital Libraries: ECDL 2001*, pages 127–136, Berlin, 2001. Springer-Verlag.

By Any Other Name: Name-based Record Disambiguation in the Social Networks and Archival Context Project

Ray R. Larson and Yiming Liu

School of Information,
University of California, Berkeley
Berkeley, California, USA, 94720-4600
{ray, yliu}@ischool.berkeley.edu

Abstract. The Social Networks and Archival Context project (SNAC) is an ongoing effort to produce software and linked data repositories connecting historical persons, families, and corporate bodies, providing information and context for humanities scholars conducting research regarding these historical entities. As SNAC aggregates archive records across multiple archival sources, effective disambiguation of entities with identical or near-identical names, and accurate merging of their records across archive boundaries, is a persistent problem. This paper describes current methodologies in record disambiguation and merging in SNAC, drawing upon results from previous experiments with supervised classification methods and a new n-gram-based approach using the Cheshire II information retrieval system.

1 Introduction

The Social Networks and Archival Context (SNAC) project is an ongoing research effort to address the challenge of improving access to primary humanities resources through the use of advanced technologies. One of the primary goals of the project is to create linked data repositories of historic persons, corporate entities, and families to existing archival descriptions, library catalogs, and authority files. To accomplish this, SNAC ingests archival records in the form of Encoded Archival Description (EAD) files and produces Encoded Archival Context - Corporate bodies, Persons, and Families (EAC-CPF) files, containing combined and interlinked representations of each entity. Archival records originate from the Library of Congress (LOC), Online Archive of California (OAC), Northwest Digital Archive (NWDA), and Virginia Heritage (VH).

Since there are multiple archival collections represented within SNAC, persons, corporate bodies, and families may have authoritative names within each individual collection, and these authoritative names may be different across collections. One particular challenge of SNAC is to identify, match, merge, and link records representing the same entities across archival collections. This also allows a far richer web of archival material and context to be presented for these historical entities, more so than each individual collection is capable of. The Virtual

International Authority File (VIAF) is the primary name authority file used in the identification, matching, and merging process, supported by supplementary name authority sources such as LOC.

The following sections of the paper describe the current state of SNAC entity merging and linking process. It first briefly reviews previous work and experimental results in entity matching and merging. Then it summarizes the current load/match/merge system for entity records, a multi-tiered architecture of matching algorithms culminating in an n-gram-driven classifier, and the results of some validation experiments based on this system. Finally, the paper outlines the limitations of the current approach and expected future work.

2 Previous work

The full scope of SNAC design and implementation, from record ingestion to frontend display, is described in detail by Larson and Janakiraman [1], and has not changed in general. In summary, XSLT transformations are applied to standard XML-based EAD archival descriptions, extracting names, dates of existence (birth and death dates, in the case of persons), occupation, subject headings, etc. These extractions are output as EAC-CPF documents, another standardized XML format. A matching and merging process is then applied to the EAC-CPF documents, using the information contained within to attempt to identify the same historical entity across all archival boundaries. Once identified, they are merged together into a single record, and any linked data records are re-established against the merged identity as appropriate. These records are then published to a dynamic Web-based frontend for user consumption.

Larson and Janakiraman experimented with various entity matching and merging processes, primarily based on entity names and existence dates. Three matching algorithms were described, beyond the trivial exact name string matching. Namely, these are:

Edit Distance – Given a name, candidate entities for merger are ranked based on the mean edit distance between the names associated with the entity and the given name. Levenshtein Distance is used as the ranking algorithm.

Shingle-based NaiveBayes – Given a name, both the query name and names of candidate entities for merger are sliced into N-character shingles. The presence or absence of a shingle is treated as a feature in a feature vector, and each candidate name is considered a class. A NaiveBayes classifier is trained on the set of candidates and asked to classify the newly given name.

Multinomial NaiveBayes – Given a name, both the query name and names of candidate entities for merger are sliced into N-character shingles. The number of occurrences of a shingle in the name is treated as a feature in a feature vector, and each candidate name is considered a class. A NaiveBayes classifier is trained on the set of candidates and asked to classify the newly given name.

Of these approaches, it was found that the Multinomial NaiveBayes algorithm produced the most accurate merger results, with 60.82% of correct target entities found as the top-ranked match based on name alone.

3 Matching System Architecture

Drawing upon the previously indicated work and some new developments via the Cheshire II IR system, the latest iteration of SNAC implements a multi-tiered matching architecture. The SNAC entity matching consists of a three-phase system.

3.1 Load

In the first `Load` phase, the system ingests EAC-CPF records and records intermediary data into a MongoDB instance, including a unique record ID, name, existence dates, and other relevant pieces of context from the EAC record. As MongoDB is a fast NoSQL/document-centric database system, this stage in essence creates a memory-backed cache for further query and merging operations.

3.2 Match

In the subsequent `Match` phase, the system conducts a multi-staged matching operation. A Cheshire II [2] installation is used to index VIAF and LOC name authority records, which are used as the basis for entity matching against EAC-CPF records. In the first stage, names are matched against each other by exact string match. As discovered by Larson and Janakiraman, many extracted names are already encoded using standard cataloging rules. As such these names contain existence dates and other detailed information, and match exactly with other extracted records and authority records using the same rules. Approximately 25 percent of the records are matched this way.

In stage 2, the Cheshire II system is queried with the entity name via an n-gram index. The details of this stage is described in Section 4. The Cheshire II system returns a ranked list of matches by n-gram overlap – that is, the name that shares the most n-grams with the query name is ranked first. Names with 80 percent overlap or more, with similar existence dates (if existence dates are available) are considered matches.

In stage 3, the previously outlined multinomial NaiveBayes method is applied to conduct matching. For performance reasons, the top 50 matches (or as many matches as possible, if less than 100) from stage 2 are processed into 3-gram shingle feature vectors, and given used as a training set to a multinomial Naive-Bayes classifier. Each entity name is similarly sliced into 3-gram shingles. The same algorithm outlined in Section 2 and [1] is used, with the top ranked result considered a match, unless even the probability estimate of the top-ranked class is still very low (this threshold being a tunable parameter).

The benefit of this system is that additional stages can be inserted at any point, with the most accurate methods executed first. In this fashion, a matching operation cascades down the stages, being returned if a match is found at any point. The more expensive matching operations do not execute until the previous, presumably less computationally expensive stages have had a chance to make the match first.

3.3 Merge

In the final `Merge` phase, records containing data belonging to matched entities are merged together to form the final EAC-CPF record, for front-end consumption. In this stage, all entities are considered de-duplicated and merged. Additional linked data, such as DBpedia information or other resources external to the library records themselves, could be inserted into the final record.

4 Ngram Indexes in Cheshire II

Placeholder

5 Validation

To validate the effectiveness of the new matching system, the experiments in Larson and Janakiraman are replicated in abbreviated form with the new algorithms. The VIAF name authority records are used as a validation dataset. The collection of 291,952 VIAF records, when including alternate names, contains 1,886,049 names. The classifier, when used, is trained on 70% of these names. For testing, 10,000 names were randomly selected out of the remaining pool. For performance reasons, the exact same setup for the experiment was run in 5 parallel processes, each with its own randomly selected 10,000 names, instead of using 50,000 names.

Results were comparable and in some cases better than the pure application of MNB. These results are outlined in Table 1.

Table 1. This table presents accuracy measures of validation experiments using ngram indexes and ngram indexes plus supervised Naive Bayes matching.

	Criteria	
Matching	First	Top 5
Ngram index only	0.7230	0.8011
Ngram index + MNB	0.7780	0.8420

6 Future work

The current SNAC entity matching infrastructure has a number of limitations. The ngram-based matching is slower than previous methods. It proved exceedingly cumbersome to replicate the full scale 70/30 cross-validation experiments as originally outlined in Larson and Janakiraman, as searching the ngram index for 10,000 records took approximately 24 hours. It may prove fruitful to improve the performance of this matching method by executing parallel lookups instead, which the current infrastructure is not set up to handle.

The system does not make full use of MongoDB's built-in facilities for query processing and MapReduce-based operations. For many intermediate stages, it would be useful to produce temporary output and use the database facilities for processing, instead of relying on application-layer code.

References

1. Larson, R.R., Janakiraman, K.: Connecting archival collections: the social networks and archival context project. In: Proceedings of the 15th international conference on Theory and practice of digital libraries: research and advanced technology for digital libraries. TPDL'11, Berlin, Heidelberg, Springer-Verlag (2011) 3–14
2. Larson, R.R., Mcdonough, J., Kuntz, L., Moon, R.: Cheshire ii: Designing a next-generation online catalog. Journal of the American Society for Information Science **47** (1996) 555–567

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

128,787 Names

All Names → Alphabetical Index →

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Featured Records

Anthony, Susan B
Berkeley Free Church
Bernstein, Leonard, 1918-
Block, Herbert, 1909-2001
Bush, Vannevar, 1890-1974
Cha, Theresa Hak Kyung
Feynman, Richard Phillips, 1918-1988
Fitzgerald, Ella, 1918-1996
Frankfurter, Felix, 1882-1965
Franklin, Benjamin, 1706-1790
Fuller, R. Buckminster (Richard Buckminster), 1895-1983
Hamilton, Alexander, 1757-1804
Luce, Clare Boothe, 1903-1987
Oppenheimer, J. Robert, 1904-1967
Patton family
Patton, George S. (George Smith), 1885-1945
Sontag, Susan, 1933-2004
Washington, George, 1732-1799

Top Occupations

- Journalists (188)
- Authors (179)
- Lawyers (172)
- Authors, American (121)
- Diplomats (115)
- Educators (96)
- Farmers (83)
- Historians (79)
- Army officers (76)
- Naval officers (76)
- Engineers (73)
- Domestics (70)
- Businessmen (69)
- Public officials (62)
- Economists (61)
- Legislators (59)
- Representatives, U.S. Congress (55)
- Jurists (53)
- Senators, U.S. Congress (52)
- Missionaries (48)
- Statesmen (48)
- Cabinet officers (46)
- Clergy (43)

Top Subjects

- World War, 1939-1945 (716)
- Montana (439)
- Idaho (328)
- World War, 1914-1918 (319)
- Education (315)
- Agriculture (301)
- Washington (State) (277)
- International relief (226)
- Colleges and Universities (215)
- Pioneers (199)
- Communism (172)
- Family (170)
- Architecture (164)
- Government and Politics (152)
- Photographs (148)
- African Americans (146)
- Japanese Americans (145)
- Refugees (141)
- Authors, American (138)
- Railroads (136)
- Music (133)
- Norwegian-Americans (128)
- Literature (127)
- Oregon (122)

Find Corporate, Personal, and Family Archival Co xt Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

128,787 Names

All Names → Alphabetical Index →

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Featured Records

Anthony, Susan B
Berkeley Free Church
Bernstein, Leonard, 1918-
Block, Herbert, 1909-2001
Bush, Vannevar, 1890-1974
Cha, Theresa Hak Kyung
Feynman, Richard Phillips, 1918-1988
Fitzgerald, Ella, 1918-1996
Frankfurter, Felix, 1882-1965
Franklin, Benjamin, 1706-1790
Fuller, R. Buckminster (Richard Buckminster), 1895-1983
Hamilton, Alexander, 1757-1804
Luce, Clare Boothe, 1903-1987
Oppenheimer, J. Robert, 1904-1967
Patton family
Patton, George S. (George Smith), 1885-1945
Sontag, Susan, 1933-2004
Washington, George, 1732-1799

Top Occupations

- Journalists (188)
- Authors (179)
- Lawyers (172)
- Authors, American (121)
- Diplomats (115)
- Educators (96)
- Farmers (83)
- Historians (79)
- Army officers (76)
- Naval officers (76)
- Engineers (73)
- Domestics (70)
- Businessmen (69)
- Public officials (62)
- Economists (61)
- Legislators (59)
- Representatives, U.S. Congress (55)
- Jurists (53)
- Senators, U.S. Congress (52)
- Missionaries (48)
- Statesmen (48)
- Cabinet officers (46)
- Clergy (43)

Top Subjects

- World War, 1939-1945 (716)
- Montana (439)
- Idaho (328)
- World War, 1914-1918 (319)
- Education (315)
- Agriculture (301)
- Washington (State) (277)
- International relief (226)
- Colleges and Universities (215)
- Pioneers (199)
- Communism (172)
- Family (170)
- Architecture (164)
- Government and Politics (152)
- Photographs (148)
- African Americans (146)
- Japanese Americans (145)
- Refugees (141)
- Authors, American (138)
- Railroads (136)
- Music (133)
- Norwegian-Americans (128)
- Literature (127)
- Oregon (122)

feedback

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

128,787 Names

All Names → Alphabetical Index →

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Featured Records

Anthony, Susan B
Berkeley Free Church
Bernstein, Leonard, 1918-
Block, Herbert, 1909-2001
Bush, Vannevar, 1890-1974
Cha, Theresa Hak Kyung
Feynman, Richard Phillips, 1918-1988
Fitzgerald, Ella, 1918-1996
Frankfurter, Felix, 1882-1965
Franklin, Benjamin, 1706-1790
Fuller, R. Buckminster (Richard Buckminster), 1895-1983
Hamilton, Alexander, 1757-1804
Luce, Clare Boothe, 1903-1987
Oppenheimer, J. Robert, 1904-1967
Patton family
Patton, George S. (George Smith), 1885-1945
Sontag, Susan, 1933-2004
Washington, George, 1732-1799

T Occupations

- [illegible] (118)
- [illegible]nals (115)
- [illegible]tors (96)
- [illegible]ers (83)
- His[illegible]ians (79)
- Army officers (76)
- Naval officers (76)
- Engineers (73)
- Domestics (70)
- Businessmen (69)
- Public officials (62)
- Economists (61)
- Legislators (59)
- Representatives, U.S. Congress (55)
- Jurists (53)
- Senators, U.S. Congress (52)
- Missionaries (48)
- Statesmen (48)
- Cabinet officers (46)
- Clergy (43)

Top Subjects

- World War, 1939-1945 (716)
- Montana (439)
- Idaho (328)
- World War, 1914-1918 (319)
- Education (315)
- Agriculture (301)
- Washington (State) (277)
- International relief (226)
- Colleges and Universities (215)
- Pioneers (199)
- Communism (172)
- Family (170)
- Architecture (164)
- Government and Politics (152)
- Photographs (148)
- African Americans (146)
- Japanese Americans (145)
- Refugees (141)
- Authors, American (138)
- Railroads (136)
- Music (133)
- Norwegian-Americans (128)
- Literature (127)
- Oregon (122)

feedback

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

128,787

All Names → Alphabetical Index →

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Featured Records

Anthony, Susan B
Berkeley Free Church
Bernstein, Leonard, 1918-
Block, Herbert, 1909-2001
Bush, Vannevar, 1890-1974
Cha, Theresa Hak Kyung
Feynman, Richard Phillips, 1918-1988
Fitzgerald, Ella, 1918-1996
Frankfurter, Felix, 1882-1965
Franklin, Benjamin, 1706-1790
Fuller, R. Buckminster (Richard Buckminster), 1895-1983
Hamilton, Alexander, 1757-1804
Luce, Clare Boothe, 1903-1987
Oppenheimer, J. Robert, 1904-1967
Patton family
Patton, George S. (George Smith), 1885-1945
Sontag, Susan, 1933-2004
Washington, George, 1732-1799

Top Occupations

- Journalists (121)
- (96)
- Historians
- Army officers
- Naval officers (76)
- Engineers (73)
- Domestics (70)
- Businessmen (69)
- Public officials (62)
- Economists (61)
- Legislators (59)
- Representatives, U.S. Congress (55)
- Jurists (53)
- Senators, U.S. Congress (52)
- Missionaries (48)
- Statesmen (48)
- Cabinet officers (46)
- Clergy (43)

Top Subjects

- World War, 1939-1945 (716)
- Montana (439)
- Idaho (328)
- World War, 1914-1918 (319)
- Education (315)
- Agriculture (301)
- Washington (State) (277)
- International relief (226)
- Colleges and Universities (215)
- Pioneers (199)
- Communism (172)
- Family (170)
- Architecture (164)
- Government and Politics (152)
- Photographs (148)
- African Americans (146)
- Japanese Americans (145)
- Refugees (141)
- Authors, American (138)
- Railroads (136)
- Music (133)
- Norwegian-Americans (128)
- Literature (127)
- Oregon (122)

feedback

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords

128,787 Names

All Names → Alphabetical Index →

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Featured Records

Anthony, Susan B
Berkeley Free Church
Bernstein, Leonard, 1918-
Block, Herbert, 1909-2001
Bush, Vannevar, 1890-1974
Cha, Theresa Hak Kyung
Feynman, Richard Phillips, 1918-1988
Fitzgerald, Ella, 1918-1996
Frankfurter, Felix, 1882-1965
Franklin, Benjamin, 1706-1790
Fuller, R. Buckminster (Richard Buckminster), 1895-1983
Hamilton, Alexander, 1757-1804
Luce, Clare Boothe, 1903-1987
Oppenheimer, J. Robert, 1904-1967
Patton family
Patton, George S. (George Smith), 1885-1945
Sontag, Susan, 1933-2004
Washington, George, 1732-1799

Top Occupations

- Journalists (188)
- Authors (179)
- Lawyers (172)
- Authors, American (121)
- Diplomats (115)
- Educators (96)
- Farmers (83)
- Historians (79)
- Army officers (76)
- Naval officers (76)
- Engineers (73)
- Domestics (70)
- Businessmen (69)
- Public officials (62)
- Economists (61)
- Legislators (59)
- Representatives, U.S. Congress (55)
- Jurists (53)
- Senators, U.S. Congress (52)
- Missionaries (48)
- Statesmen (48)
- Cabinet officers (46)
- Clergy (43)

Top Subjects

- World War, 1939-1945 (716)
- Montana (439)
- Idaho (328)
- World War, 1914-1918 (319)
- Education (315)
- Agriculture (301)
- Washington (State) (277)
- International relief (226)
- Colleges and Universities (215)
- Pioneers (199)
- Communism (172)
- Family (170)
- Architecture (164)
- Government and Politics (152)
- Photographs (148)
- African Americans (146)
- Japanese Americans (145)
- Refugees (141)
- Authors, American (138)
- Railroads (136)
- Music (133)
- Norwegian-Americans (128)
- Literature (127)
- Oregon (122)

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

95,582 Person Names

Person Names → Alphabetical Index → A

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A, Charles.
A, Francis.
A, Hert.
A, Joseph.
A, Kulling F.
A, Nelson.
A, Rhasha.
A, Robert.
A.
A. B. Hansen, Jr.
A. Burton Clarke.
A. Carnegie Ross.
A. Cooke Garnett.
A. D. C., editor.
A. D. Peters, Literary.
A. E. Dick Howard.
A. Grove Day.
A. Iu. Bykov.
A. Lamont Smith.
A. Linwood (Abner Linwood).

Top Occupations

- Journalists (185)
- Authors (176)
- Lawyers (165)
- Authors, American (118)
- Diplomats (113)
- Educators (94)
- Farmers (82)
- Historians (77)
- Army officers (75)
- Naval officers (74)
- Engineers (71)
- Domestics (70)
- Businessmen (60)
- Public officials (60)
- Economists (59)
- Legislators (54)
- Representatives, U.S. Congress (54)
- Jurists (52)
- Senators, U.S. Congress (50)
- Missionaries (47)
- Statesmen (47)
- Cabinet officers (43)
- Photographers (43)

Top Subjects

- World War, 1939-1945 (579)
- World War, 1914-1918 (269)
- Education (243)
- International relief (197)
- Montana (193)
- Washington (State) (193)
- Agriculture (176)
- Family (167)
- Pioneers (158)
- Communism (152)
- Idaho (151)
- Authors, American (132)
- Colleges and Universities (131)
- Architecture (130)
- Norwegian-Americans (128)
- Photographs (115)
- Japanese Americans (110)
- Christmas (103)
- Literature (103)
- Refugees (101)
- Music (100)
- Ocean travel (95)
- Presidents (95)
- Railroads (91)

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

31,287 Corporate Body Names

Corporate Body Names → Alphabetical Index → A

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A & D Software.
A & M Records (Firm)
A Child's Garden Press.
A Contemporary Theatre (Seattle, Wash.).
A Different Light Bookstore.
A Gift from the estate of J. Harold Montague.
A Quaker Action Group.
A. & G.
A. & G. Carisch & C.
A. & G. Carisch & Co.
A. & G. Price.
A. & S. Nordheimer.
A. & W. Geib.
A. A. Rosenberg.
A. A. Van Gelder.
A. Albaret, Dakar.
A. Allmuth.
A. Arevalo, Libreria Espanola y Extrajera, Tangiers.
A. B. Dollar (steam launch).
A. Bock.

Top Occupations

- Clergy (17)
- Publishers and publishing (5)
- (4)
- Booksellers and bookselling (4)
- Fire fighters (4)
- Physicians (4)
- Television producers and directors (4)
- Architects (3)
- Mayors (3)
- Military attachés (3)
- Antiquarian booksellers (2)
- Authors, American (2)
- Economists (2)
- Journalists (2)
- Nurses (2)
- Orange growers (2)
- Women fire fighters (2)
- . 7$aPhotoprints (1)
- Activists (1)
- Aeronautical engineers (1)
- Air pilots (1)
- Art dealers (1)
- Artists (1)

Top Subjects

- Montana (240)
- Idaho (173)
- World War, 1939-1945 (124)
- Agriculture (111)
- Colleges and Universities (80)
- Government and Politics (76)
- Washington (State) (75)
- African Americans (70)
- Education (69)
- Business, Industry, and Labor (63)
- Clubs (51)
- College students (49)
- World War, 1914-1918 (45)
- Transportation (42)
- Railroads (40)
- Refugees (38)
- Schools (35)
- Architecture (30)
- International relief (29)
- Oregon (29)
- Photographs (29)
- Music (28)
- National socialism (28)
- Japanese Americans (26)

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords search

1,918 Family Names

Family Names → Alphabetical Index → A

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Aadneram family
Aamot family
Aas family
Aase family
Aaslian family
Abbott family
Abelsen family
Abelson family
Abernathy family
Abiko family
Abrahamsson family
Accornero family
Adair family
Adams family
Adams Postcards Historical Note, 1892-2007
Adams, Linda Friend, 1943-2008
Adamson family
Addington family
Adee family
Adkins family

Top Occupations

- Businessmen (8)
- Lawyers (6)
- Legislators (5)
- Physicians (4)
- Ranchers (4)
- Authors, English (3)
- Cabinet officers (3)
- Judges (3)
- Authors (2)
- Civic leaders (2)
- Diplomats (2)
- Governors (2)
- Historians (2)
- Naval officers (2)
- Periodical editors (2)
- Publishers and publishing (2)
- Senators, U.S. Congress (2)
- Sheep ranchers (2)
- Soldiers (2)
- Suffragists (2)
- (1)
- Abolitionists (1)
- Ambassadors (1)
- Ambassadors' spouses (1)

Top Subjects

- Pioneers (25)
- Agriculture (14)
- World War, 1939-1945 (13)
- Home and Family (10)
- Japanese Americans (9)
- Washington (State) (9)
- Japanese American families (8)
- Slavery (8)
- Oregon (7)
- Schools (7)
- Montana (6)
- Practice of law (6)
- Ranches (6)
- Ranching (6)
- African Americans (5)
- City and Town Life (5)
- Music (5)
- Railroads (5)
- Real estate development (5)
- Russians United States (5)
- World War, 1914-1918 (5)
- Architecture (4)
- Baptists (4)
- Businessmen (4)

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

enter a name or keywords | search | *advanced...* limit to section | cpfDescription

- eac-cpf
- control
- cpfDescription ✓
- identity
- description
- existDates
- biogHist
- generalContext
- occupation
- localDescription
- relations

31,287 Corporate Body Names

Corporate Body Names → Alphabetical Index → A

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

A & D Software.
A & M Records (Firm)
A Child's Garden Press.
A Contemporary Theatre (Seattle, Wash.).
A Different Light Bookstore.
A Gift from the estate of J. Harold Montague.
A Quaker Action Group.
A. & G.
A. & G. Carisch & C.
A. & G. Carisch & Co.
A. & G. Price.
A. & S. Nordheimer.
A. & W. Geib.
A. A. Rosenberg.
A. A. Van Gelder.
A. Albaret, Dakar.
A. Allmuth.
A. Arevalo, Libreria Espanola y Extrajera, Tangiers.
A. B. Dollar (steam launch).
A. Bock.

Subjects

ntana (240)
ho (173)
rld War, 1939-1945 (124)
culture (111)
leges and Universities (80)
vernment and Politics (76)
shington (State) (75)
can Americans (70)
cation (69)
siness, Industry, and Labor (63)
os (51)
lege students (49)
rld War, 1914-1918 (45)
nsportation (42)
roads (40)
ugees (38)
ools (35)
hitecture (30)
rnational relief (29)
gon (29)
tographs (29)
sic (28)
tional socialism (28)
Japanese Americans (26)

feedback

Advanced limits match EAC sections

Advanced EAC Search

https://spreadsheets.google.com/pub?key=0AmlczcEUr8GxdGRjeEJ6S2tBU0lmMjlDaWZZdFNiQmc&h

Google

Advanced EAC Search : Sheet1

eac-cpf	↴			entire record
	control			maintenance and sources
	cpfDescription	↴		**default search**
		identity		contains the name or names used by the entity over the course of the entity's existence.
		description	↴	accommodates a variety of both controlled and prose descriptions of entities, closely reflecting the descriptive categories outlined in ISAAR (CPF).
			existDates	dates of existence of the entity being described.
			biogHist	Includes discursive text providing biographical and/or historical information about the entity being described.
			generalContext	Includes information about the general social and cultural context of the entity being described.
			occupation	Includes relevant occupations held by the entity being described.
			localDescription	An element intended to extend the descriptive categories available in a local system. [Subjects]
		relations	↴	Related Records
			cpfRelation	Related Named Entities
			resourceRelation	Related Resources

Published by Google Docs – Report Abuse – Updated automatically every 5 minutes

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

Robert op

search

advanced... limit to section cpfDescription

Oppenheimer, J. Robert, 1904-1967.

128,787 Names

All Names → Alphabetical Index →

0 A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Top Occupations

- Journalists (188)
- Authors (179)
- Lawyers (172)
- Authors, American (121)
- Diplomats (115)
- Educators (96)

Top Subjects

- World War, 1939-1945 (716)
- Montana (439)
- Idaho (328)
- World War, 1914-1918 (319)
- Education (315)
- Agriculture (301)

Find Corporate, Personal, and Family Archival Context Records

SNAC prototype

All | Person | Corporate Body | Family

Oppenheimer, J. Robert, 1904-1 search

narrow results by

Occupation

- Physicists (8)
- Educators (6)
- Public officials (3)
- Atomic energy commissioners (1)
- Chemists (1)
- Dramatists (1)
- Engineers (1)
- Jurists (1)
- Librarians of Congress (1)
- Mathematicians (1)
- Oceanographers (1)
- Poets (1)
- Science administrators (1)
- Scientists (1)

Subject

- Science (7)
- World War, 1939-1945 (7)
- Nuclear energy (6)
- Physics (6)
- Atomic bomb (4)
- Nuclear physics (4)

69 EAC-CPF Records

Oppenheimer, J. Robert, 1904-1967.
...person Oppenheimer, J. Robert, 1904-1967. DLC Oppenheimer,...
...1904-1967 VIAF Oppenheimer, J. Robert, 1904-1967 VIAF...
...OAC Oppenheimer, J. Robert (Julius Robert), 1904-1967 VIAF...

Yukawa, Hideki, 1907-1982.
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Federation of American Scientists.
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Garrison, Lloyd K. (Lloyd Kirkham), b. 1897.
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Manley, John Henry, 1907-
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Marks, Herbert S., 1907-1960.
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Christy, Robert F.
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Kusaka, Shuichi, 1915-
...Oppenheimer, J. Robert, 1904-1967. recordId: DLC.ms998007.c...

Serber, R. (Robert).

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

▾ Archival Collections (15)

creatorOf (1) | referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

▸ People (63)
▸ Corporate Bodies (7)
▸ Resources (85)
▸ Linked Data (1)

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

Archival Collections (15)

creatorOf (1) referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

People (63)

Corporate Bodies (7)

Resources (85)

Linked Data (1)

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Bi

1904, Apr
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

Archival Collections (15)

creatorOf (1) referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

People (63)

Corporate Bodies (7)

Resources (85)

Linked Data (1)

feedback

Find Corporate, Personal, and Fami

alternative forms of name

Oppenheimer, J. Robert.
Oppenheimer, J. Robert (Julius Robert), 1904-1967
Oppenheimer, Robert, 1904-1967
Oppenheimer, J. Robert, 1904-1967
Oppenheimer, Julius Robert, 1904-1967
Oppenheimer, Jacob Robert, 1904-1967
Oppenheimer, J. Robert 1904-1967
Ou-pên-hai-mo, 1904-1967
Oppenheimer, Robert
Oppenheimer, Robert 1904-1967
Ou-pên-hai-mo, 1904-1967

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert

(1904, Apr. 22 - 1967, Feb. 18) Male

--› Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographic

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

ries

tions (15)

creatorOf (1) referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

People (63)
Corporate Bodies (7)
Resources (85)
Linked Data (1)

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

190 Apr. 22
w York, N.Y.

Mass.
arvard University, Cambridge, Mass.

192
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

▾ Archival Collections (15)

creatorOf (1) referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

▸ People (63)
▸ Corporate Bodies (7)
▸ Resources (85)
▸ Linked Data (1)

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.
1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.
1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England
1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany
1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland
1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.
1940
Married Katherine Harrison
1943-1945
Los Alamos, N. Mex.

Related Entries

Archival Collections (15)

creatorOf (1) referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

People (63)

Corporate Bodies (7)

Resources (85)

Linked Data (1)

feedback

- international affairs.
- Science and state.
- Science Study and teaching.
- Science--History--20th century.
- Science--Moral and ethical aspects.
- Science--Security measures.
- Science--Social aspects.
- Science--Societies, etc.
- Security clearances--United States.
- World War, 1939-1945--Science.

Princeton, N.J.
Director and professor of physics, Institute for Advanced Study, Princeton, N.J.

1949-1955
Member, Board of Overseers of Harvard College

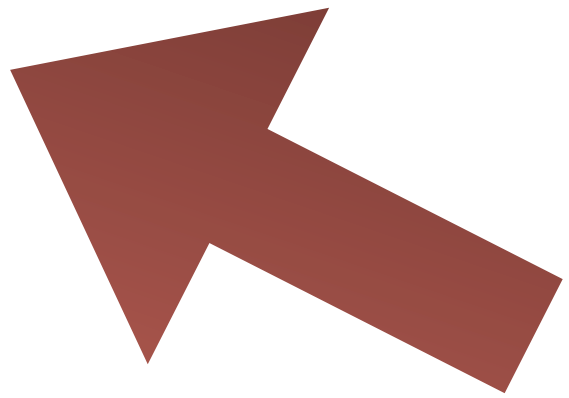
1954
(1) Denied security clearance by the Personnel Security Board of the U.S. Atomic Energy Commission (2) Published Science and the Common Understanding (New York: Simon and Schuster. 120 pp.)

1955
Published The Open Mind (New York: Simon and Schuster. 146 pp.)

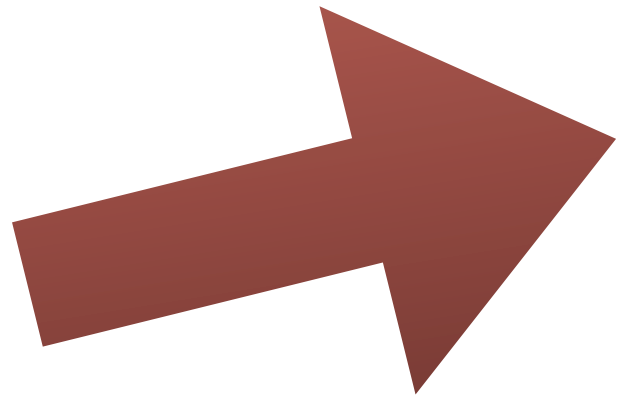
1963
Received the Enrico Fermi Award

1967, Feb. 18
Princeton, N.J.
Died, Princeton, N.J.

From the finding aid for J. Robert Oppenheimer Papers 1799-1980 (bulk 1947-1967) (Manuscript Division Library of Congress)



Find Corporate, Personal, and Family Archival Context Records



radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

▾ Archival Collections (15)

creatorOf (1) | referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

▸ People (63)
▸ Corporate Bodies (7)
▸ Resources (85)
▸ Linked Data (1)

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United Stat* ...ish

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

▾ Archival Collections (15)

creatorOf (1) | referencedIn (14)

J. Robert Oppenheimer Papers, 1799-1980 (bulk 1947-1967) ↗
Manuscript Division Library of Congress

▸ People (63)
▸ Corporate Bodies (7)
▸ Resources (85)
▸ Linked Data (1)

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

▾ Archival Collections (15)

creatorOf (1) | referencedIn (14)

Ava Helen and Linus Pauling Papers 1910-1994 (bulk 1922-1991) ↗
Oregon State University The Valley Library, Special Collections

Carl Eckart Papers, 1921-1973 (bulk 1935-1970) ↗
Manuscript Division Library of Congress

Charles Christian Lauritsen papers, 1927-1977 ↗
California Institute of Technology. Archives.

Denise Royal Papers 1967-1970 ↗
University of Oregon LibrariesSpecial Collections & University Archives

Glenn Theodore Seaborg Papers, 1866-1999 (bulk 1940-

▸ People (63)
▸ Corporate Bodies (7)
▸ Resources (85)
▸ Linked Data (1)

feedback

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

▸ Archival Collections (15)

▾ People (63)

Alsop, Joseph, 1910-1989. associatedWith
Alsop, Stewart. associatedWith
Bacher, Robert F. (Robert Fox). associatedWith
Bethe, Hans A. (Hans Albrecht), 1906-2005. correspondedWith
Birge, Raymond T. (Raymond Thayer), b. 1887. correspondedWith
Bloch, Felix, 1905- correspondedWith
Bohr, Niels, 1885-1962. associatedWith
Born, Max, 1882-1970. correspondedWith
Boyd, Julian P. (Julian Parks), 1903-1980. correspondedWith

▸ Corporate Bodies (7)

▸ Resources (85)

▸ Linked Data (1)

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

- Archival Collections (15)
- People (63)
- Corporate Bodies (7)
  - Federation of American Scientists. associatedWith
  - Institute for Advanced Study (Princeton, N.J.) associatedWith
  - Los Alamos Scientific Laboratory. associatedWith
  - National Academy of Sciences (U.S.) associatedWith
  - Twentieth Century Fund. associatedWith
  - U.S. Atomic Energy Commission. associatedWith
  - Unesco. associatedWith
- Resources (85)
- Linked Data (1)

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

1943-1945
Los Alamos, N. Mex.

Related Entries

- Archival Collections (15)
- People (63)
- Corporate Bodies (7)
- Resources (85)
  - A Report on the international control of atomic energy creatorOf ↗
  - A conversation with physicists Robert Oppenheimer and Niels Bohr creatorOf ↗
  - Address at the National Book awards ceremony creatorOf ↗
  - Atom and void essays on science and community creatorOf ↗
  - Atomkraft und menschliche Freiheit creatorOf ↗
  - Baut den Frieden Wir alle sind verantwortlich creatorOf ↗
  - Ben Franklin lecture at the University of Pennsylvania. creatorOf ↗
- Linked Data (1)

Search results for 'ti:A conversation with physicists Robert Oppenheimer and Niels Bohr au:Oppenheimer, J. Robert, 1904-1967.'

Results 1-1 of about 1 (.16 seconds) « First ‹ Prev 1 Next ›

Select All Clear All Save to: [New List] Save Sort by: Relevance Save Search

1. A conversation with physicists Robert Oppenheimer and Niels Bohr
by J Robert Oppenheimer; Niels Bohr; Oppenheimer (J. Robert) Collection (Library of Congress)
Music : Reel-to-reel tape
Language: English
Publisher: 1958.

Results 1-1 of about 1 (.16 seconds) « First ‹ Prev 1 Next ›

Find Corporate, Personal, and Family Archival Context Records

radial graph demo
view source EAC-CPF

Oppenheimer, J. Robert, 1904-1967. DLC

(1904, Apr. 22 - 1967, Feb. 18) Male *United States* English

⇢ Alternative forms of name

Occupations

- Educators.
- Physicists.

Subjects

- Atomic bomb.
- Exchange of publications.
- Humanitarianism.
- Internal security--United States.
- Loyalty.
- Nuclear disarmament.
- Nuclear energy--Research.

Biographical History

1904, Apr. 22
New York, N.Y.
Born, New York, N.Y.

1925
Cambridge, Mass.
A.B., Harvard University, Cambridge, Mass.

1925-1926
Cambridge, England
Student, University of Cambridge, Cambridge, England

1927
Göttingen, Germany
Ph. D., University of Göttingen, Göttingen, Germany

1928-1929
Fellow, International Education Board,Leiden University, Leiden, Netherlands, and University of Zurich, Zurich, Switzerland

1929-1947
Pasadena, Calif.
Assistant professor, associate professor, and professor of physics at the University of California, Berkeley, and California Institute of Technology, Pasadena, Calif.

1940
Married Katherine Harrison

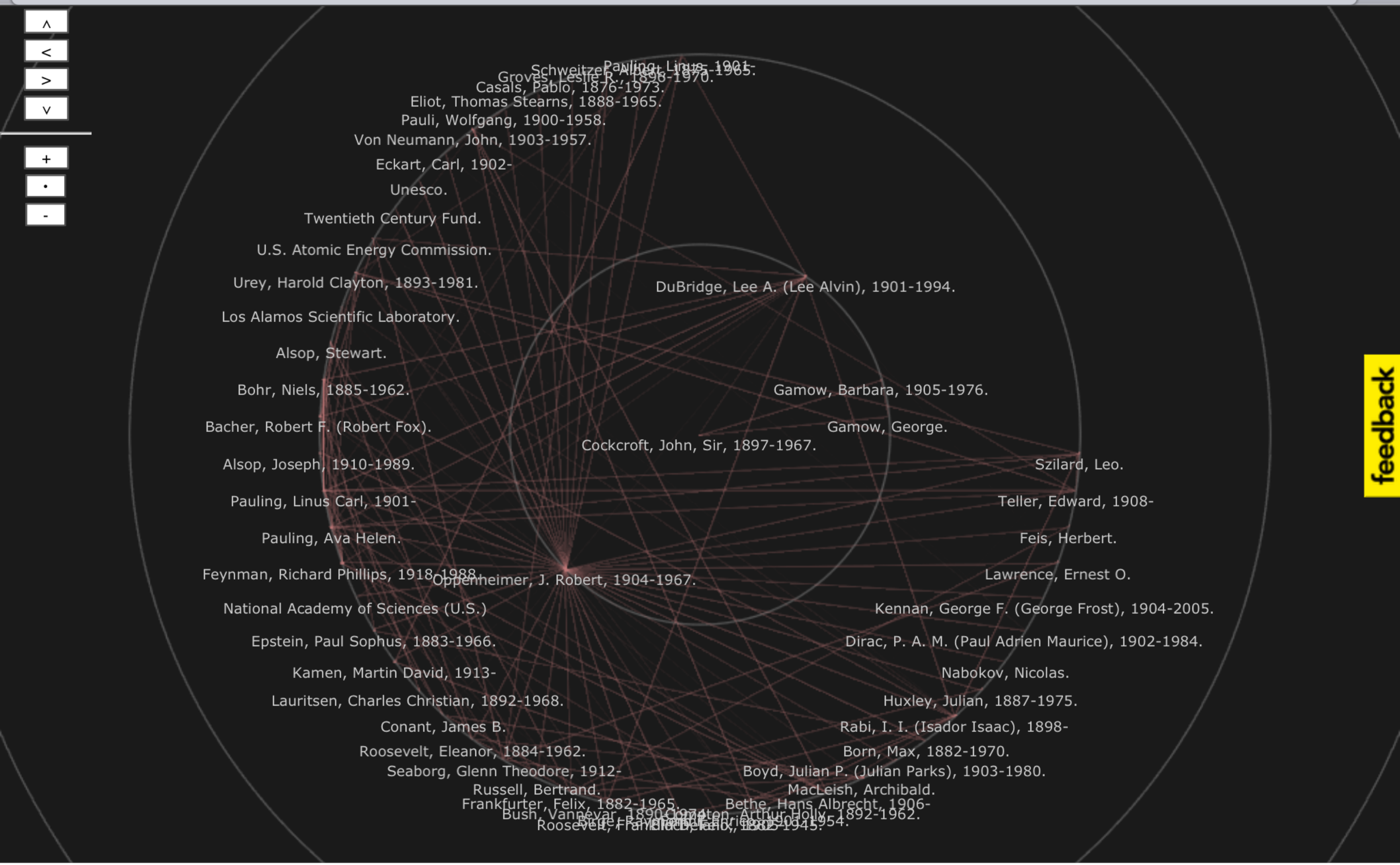
1943-1945
Los Alamos, N. Mex.

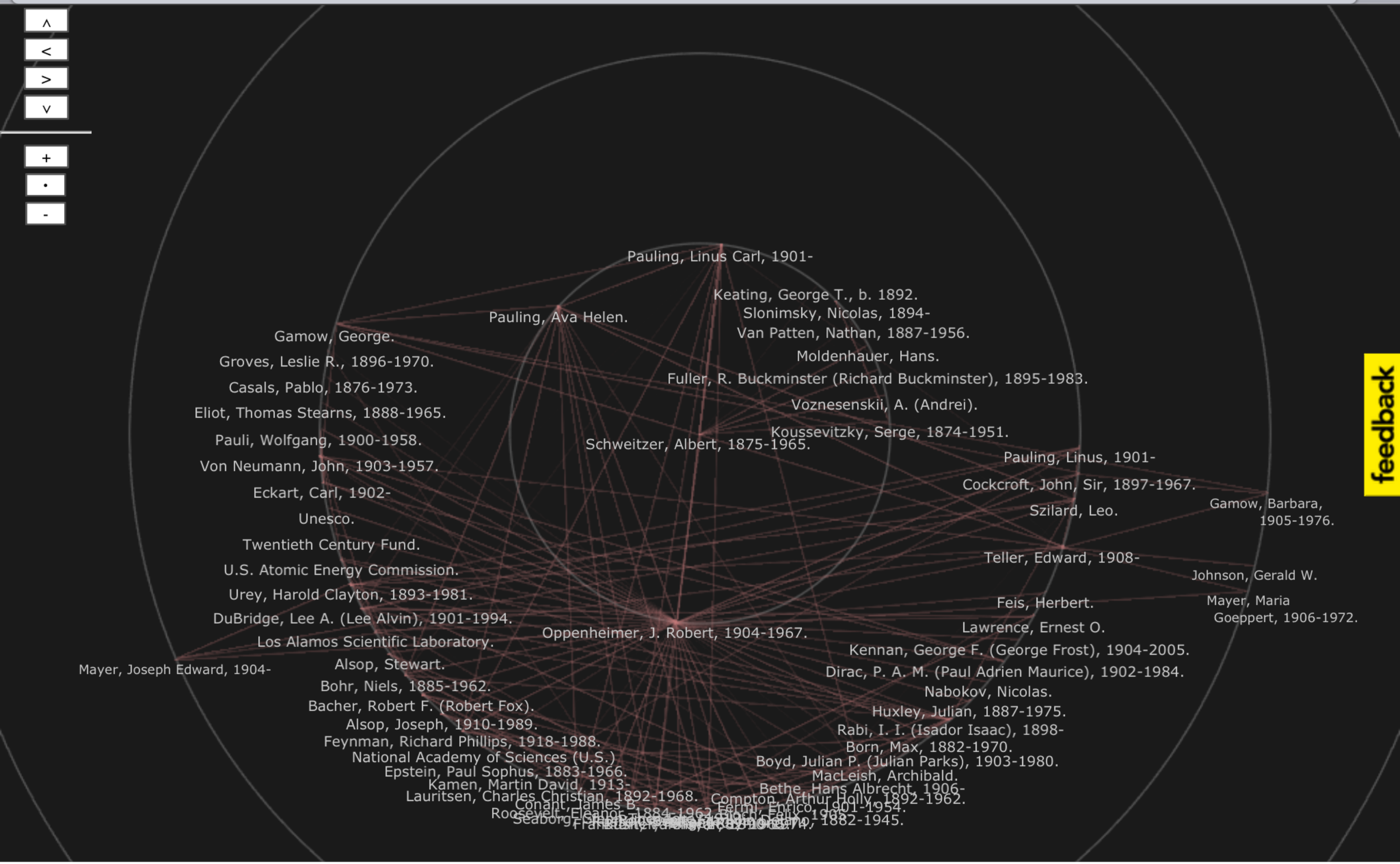
Related Entries

- Archival Collections (15)
- People (63)
- Corporate Bodies (7)
- Resources (85)
- Linked Data (1)

sameAs
http://viaf.org/viaf/44376228 ↗

feedback





Guide to the Robert Creeley | Screenr – Record

www.oac.cdlib.org/findaid/ark:/13030/tf7b69n911/?query=creeley

OAC
Online Archive of California

Search OAC
go

Home | Browse Institutions | Browse Collections | Browse Map | About OAC | Help | What is OAC?

> Home > Stanford University > Manuscripts Division

Share / Save

Collection Guide

http://www.oac.cdlib.org/findaid/ark:/13030/tf7b69n911

Collection Title: Guide to the Robert Creeley Papers, 1950-1997 M0662

Collection Number: M0662

Get Items: No online items

Contact Stanford University::Manuscripts Division

View entire collection guide

PDF (0.76 Mb) HTML [1187 hits]

Search this collection

creeley
go
Entire Collection Guide | Online Items

Collection Overview

Description The Robert Creeley Papers document the life work of a leading American poet of the 20th century, one of the core members of the "Black Mountain School." They also document several important movements in American poetics in the second half of the century. The papers include Creeley's personal and professional correspondence, journals, business records, personal mementos, clippings, artwork, and other documents generated and collected by him from 1950 to 1997.

Background Recognized as a seminal figure of American letters in the second half of the 20th

Table of contents

1187 Search hits Clear search hits

Collection Overview
Collection Details [90 hits]
Biographical chronology
Processing Information note
Scope and Content
Biographical note
Preferred Citation
Access Restrictions

context

Guide to the Robert Creeley | Screenr – Record

www.oac.cdlib.org/findaid/ark:/13030/tf7b69n911/?query=creeley

OAC
Online Archive of California

Home | Browse Institutions | Browse Collections | Browse Map | About OA

> Home > Stanford University > Manuscripts Division

Collection Guide

Collection Title: Guide to the Robert Creeley Papers, 1950-1997 M06

Collection Number: M0662

Get Items: No online items

Contact Stanford University::Manuscripts Division

Collection Overview

Description The Robert **Creeley** Papers document the life work of a leading Ameri
of the 20th century, one of the core members of the "Black Mountain S
They also document several important movements in American poetics
second half of the century. The papers include **Creeley's** personal and
professional correspondence, journals, business records, personal mer
clippings, artwork, and other documents generated and collected by hi
1950 to 1997.

Background Recognized as a seminal figure of American letters in the second half

archival context network

Creeley, Robert.
Ginsberg, Allen, 1926-1997.
Snyder, Gary.
O, John.
Howe, Fanny Quincy.
Brautigan, Richard.
Allen, Donald Merriam.
Taggart, John, 1942-
Hyman, Stanley Edgar, 1919-1970.
Levertov,Denise.
Whalen, Phil.
Ignatow, David, 1914-
Mayer, Bernadette.
Blackburn, Paul.
Degnan, June Oppen.
Koch, Peter Rutledge.
United Artists (New York, N.Y.)
Kyger, Joanne.
Bernstein, Charles, 1950-
Loewinsohn, Ron.
Reznikoff, Charles, 1894-1976.
Bromige, David.
Archive for New Poetry (University of California, San Diego)..
Ellingham, Lewis.
Vermont, Charlie.
Frym, Gloria.
Fraser, Kathleen, 1935-

Guide to the History of Oregon State University Oral Histories and Sound Recordings 1956-1980

Print this Finding Aid | Email this Finding Aid

Search

Search

Results +/-

Table of Contents

Overview of the Collection

Historical Note

Content Description

+/- Use of the Collection

Overview of the Collection +/-

Creator: Van Loan, Lillian Schroeder.

Title: History of Oregon State University Oral Histories and Sound Recordings

Dates: 1956-1980 (inclusive)
1956-1969 (bulk)

Quantity: 0.9 cubic foot, including 22 reel-to-reel audiotapes and 11 audiocassettes (3 boxes)

Collection Number: OH 03

Summary: The History of Oregon State University Oral Histories and Sound Recordings consist predominantly of oral histories conducted in 1956 by Lillian Van Loan for her doctoral dissertation on the history of Oregon State College and subsequent interviews done by Ilona Fry in 1980 on the development of liberal arts at Oregon State University.

Repository: Oregon State University Libraries

Stanley Edgar Hyman papers, 1932-1978

Some or all content stored offsite.

Collection Summary

Title	Stanley Edgar Hyman papers, 1932-1978
Span Dates	1932-1978
Bulk Dates	(bulk 1938-1970)
ID No.	MSS58941
Creator	Hyman, Stanley Edgar, 1919-1970
Extent	14,000 items ; 47 containers ; 18.6 linear feet
Language	Collection material in English
Location	Manuscript Division, Library of Congress, Washington, D.C.
Summary	Literary critic and educator. Correspondence, memoranda, journal, manuscripts of articles, book reviews, and books, research material, notes, reports, and other papers relating to Hyman's career as literary critic, book reviewer, and professor of language, literature, and the history of myth and ritual at Bennington College, Bennington, Ve... special interest are files pertaining to his book review column published in the *New Leader* and letters writt... his wife, Shirley Jackson, and by his friend and mentor, Kenneth Burke.

gist: 1593245 download fork

Description: SNAC RDF Snippets
Public Clone URL: git://gist.github.com/1593245.git
Embed All Files: show embed

DATA.md # embed raw

SNAC RDF Snippets

This gist gives you small samples from the large XML data files found inside 20110815-graphML+rdf.tar.gz:

- 80M `eac.rdf` sample excerpt
- 1.5M `snac-viaf.rdf` sample excerpt

Related Graph Database

- 155M `graphML.xml` sample excerpt

SPARQL Endpoint

http://socialarchive.iath.virginia.edu/sparql/

Related Collections API

Coming soon

tingletech (owner)

Revisions

f0e217 tingletech 4 months ago
871087 tingletech 4 months ago
48c865 tingletech 4 months ago
643ad1 tingletech 4 months ago
cafaeb tingletech 4 months ago
01da26 tingletech 4 months ago
d4e271 tingletech 4 months ago
fe1307 tingletech 4 months ago
ae22b3 tingletech 4 months ago
849985 tingletech 4 months ago